Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 36, NUMBER 10 OCTOBER, 1946

Cesium Vapor Lamps*

N. C. BEESE
Research Department, Westinghouse Lamp Division, Bloomfield, New Jersey
(Received August 14, 1946)

A study of the characteristics of low voltage arc discharges in cesium vapor led to the development of a practical lamp that was used in a successful communication system using infra-red radiation. The radiated energy is confined almost entirely to the cesium resonance lines at 8521 and 8944 angstroms. Data are presented that give structural details, operating characteristics, infra-red output, and efficiencies of several lamps of different sizes. Excellent modulability is attained throughout the audiofrequency range with current modulation up to 100 percent.

A CESIUM vapor lamp development program was undertaken in the engineering and research laboratories of the Westinghouse Lamp Division as a wartime project. The work was sponsored by the Navy Department, Bureau of Ships. The desired end result was a source that had an intensity at least equal to that of a 200-watt projection lamp as measured with a cesium photo-cell through filters that gave equal visual range to the two lamps, was capable of being modulated with audiofrequencies between 200 and 3000 cycles per second, was of such size as to produce a 25° beam when placed in a 15-inch reflector, and had sufficient ruggedness to withstand shipboard use with at least 20 hours useful life. Another objective was to design a lamp of 500 watts power with 360-degree horizontal and 60-degree vertical beam spread.

Exploratory work starting in April, 1944 was done with lamps of various sizes and shapes to investigate factors such as current, voltage, bulb temperature which regulates cesium vapor pressure, kind and amount of gas filling, type and size of electrodes, envelope glass with suitable protective coating to resist chemical attack by the cesium vapor, and infra-red output and life. As quite frequently happens a very simple design was the most practical. Laboratory experimental lamps were burned in Dewar vacuum flasks of small size to conserve heat and generate a sufficiently high pressure of cesium vapor. Commercial lamps were mounted inside an evacuated outer envelope. The infra-red output from these lamps was measured in relative terms with a cesium photo-cell, high sensitivity galvanometer, and suitable filters. This was a simple and quick method to note changes in infra-red output and efficiency with modifications in lamp design. The output in absolute units was measured with a calibrated thermopile and appropriate filters.

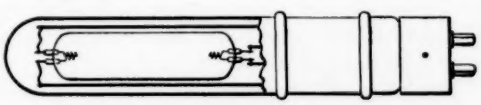


FIG. 1. Cesium vapor lamp.

* Publication of this manuscript was approved by the Navy Department, Security Review Section.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

TABLE I. Radiant energy from an experimental lamp burned at 5 amperes, mounted in a Dewar flask.

Wave-lengths	Relative intensity
10124A	5.0
10026	3.0
8943	75.0
8521	100.0
8079	1.4
7944	1.2
7609	0.7
6973	2.9
6723	1.9
6215	1.6
4593	1.4
4555	2.0

100-WATT CESIUM VAPOR LAMP

The lamp that was adopted for infra-red communication had a structure shown in Fig. 1. The source of light was a tubular bulb 1.37 inches in diameter about 5 inches long, with a spacing of 3 inches between electrodes. These dimensions gave the required beam spread in a convenient size aluminum reflector. The bulb is coated internally with a special glaze to resist the attack of ionized cesium vapor. It is fitted with two helical tungsten electrodes of small coil dimensions to anchor the arc to the center of the lamp. One or both of these may be coated with barium and strontium oxides. The outer envelope is 2 inches in diameter and has two ridges, the band between them serving to support the lamp and to maintain accurate alignment in relationship to the focal point of the reflector. A four-prong plastic base is cemented to the lamp which fits into a light weight plastic socket. The clamping band is at the center of gravity of this assembly when the lamp is in its normal horizontal position.

A gas filling of 20 to 22 cm of argon produced high efficiency in the cesium resonance lines. Lower gas pressures gave lower infra-red output while higher pressures caused increased visible light and instability. Argon gas was selected for its availability and low cost. Krypton is a heavier gas and increases the useful infra-red about 10 to 15 percent in this size lamp. Several tenths of a gram of pure cesium metal is distilled into the lamp to form the active agent in producing near infra-red radiations.

The rating of 5.5 amperes 18 volts was selected for long life with good efficiency and high infra-red output. Life tests showed that these lamps burned about 500 hours at this rating with either a.c. or d.c., and burned equally well in the horizontal or vertical positions. With audiofrequency modulation superimposed upon the normal d.c. current, the increased wattage consumed by the lamp causes a decrease in life. A forced life test with 100 percent current modulation at 1000 cycles with continuous burning gave 50 to 150 hours life. End of life is determined by a sudden decrease of infra-red energy because of loss of cesium vapor that has been absorbed by the glass bulb. The life can be regulated by the amount of cesium introduced into the lamp, the excellence of glaze coating and bulb temperature which is dependent upon the lamp wattage.

A standard inner bulb was mounted in a double wall Dewar flask with a thermocouple junction attached to it with a small amount of kaolin and water glass paste. This should give a rather close approximation to the temperature of the cesium metal at the bottom of the lamp, but will give a value that is somewhat lower than that attained in an evacuated outer housing. For 100 watts power consumption the temperature of the cesium was approximately 300°C which corresponds to a vapor pressure of about 2 mm of mercury.

MEASUREMENT OF INFRA-RED RADIATION

Cesium vapor lamps, when operated under suitable conditions, generate near infra-red radiation with high efficiency. Most of this radiation is confined to the resonance radiation at 8521A and 8944A. Spectrograms taken on infra-red sensitive film show these two lines very strong while the other lines, including the visible, are relatively weak. The spectrograms showed no reversal in the resonance lines at the normal operating current. Measurements with a Beckman spectrophotometer, recorded in Table I, showed that practically all the radiated energy between 10124A and 4555A was confined to the two resonance lines. Krefft[1] measured the intensity of the spectral lines emitted by cesium vapor arcs over a pressure range of 0.1 mm to 28 mm. At 4-mm pressure both resonance lines are strong emission lines while at 20-mm pres-

[1] H. Krefft, Zeits. f. Physik **77**, 752 (1932).

sure these two lines show strong reversal. Mohler[2] and Beutell[3] experimented with low pressure cesium vapor discharges and measured some of their characteristics which indicated that appreciable energy was radiated in the near infra-red region. At relatively high cesium vapor pressures and current densities Mohler[4] calculated high luminous efficiency for the continuous spectrum that was found under those conditions.

The infra-red radiations from a cesium vapor lamp are readily measured with a cesium photocell since it has a high response to its own radiations. A Wratten No. 87 filter is satisfactory to eliminate visible radiation from the cesium vapor lamp being measured or from the tungsten filament lamp that might be used as a standard of comparison. A Corning No. 254 glass filter 2.3 mm thick was also used with the standard 500-watt tungsten filament lamp. This filter and lamp combination gave a visual security equal to the Wratten No. 87 filter when used with the cesium lamps. With the particular RCA No. 919 photo-cell used, a normal cesium vapor lamp operating at equilibrium with 5.5 amperes had a rated output of 500 to 700 equivalent watts of tungsten radiation with equal security. This output index was dependent upon the spectral sensitivity of the photo-cell so could not be used for inter-laboratory comparisons, but served very well to detect changes in infra-red output from our experimental lamps as modifications were made in lamp structures.

A thermopile calibrated with a Bureau of Standards standard lamp, a high sensitivity galvanometer and several filters formed a measuring device that gave results in absolute units. The following table of data, Table II, shows the characteristics of an average commercial CL-2 lamp (Westinghouse designation for this lamp as made for the Navy) burned on direct current.

The efficiency of a lamp is dependent upon its arc length and condition of burning as well as purity of gas and other factors inherent in its structure. A lamp similar to the CL-2 lamp but with 4.25 inches arc length burned in double wall Dewar flask at 5.5 amperes 17.5 volts gave 21.4 watts or 22.2 percent of the total wattage supplied to the lamp in the near infra-red. When mounted in an evacuated single wall outer envelope this lamp operated at the same current and voltage but produced 24.1 watts or 25 percent of its applied wattage in the infra-red. The efficiency increased with length of arc stream, other factors remaining the same, since the electrode losses become a smaller part of the total wattage of the lamp.

Lamps 1.66 inches in diameter, with arc lengths of 4, 8, and 12 inches had the same infra-red output per unit of length when burned in a horizontal position at 5.5 amperes and showed an electrode loss of 9 volts and a gradient of 1.25 volts per cm. Using data from 2-inch diameter lamps mounted in vacuum outer jackets an electrode loss of 7 volts and a gradient of 0.97 volt per cm were obtained when the lamps were burned in a vertical position on d.c. These data were for lamps filled with 20-cm argon gas. Changes in gas pressure and kind of inert gas used as well as modifications in lamp structure will change these values.

EFFECT OF GAS IMPURITIES

In the course of our investigations it became desirable to learn the effects of small percentages of impurities such as oxygen, nitrogen, and hydrogen in the normally pure rare gas filling. If some or all produced deleterious effects, great effort should be expended in trying to remove them, but if a beneficial effect is produced by their addition, steps should be taken to introduce the optimum amounts. The experimental work consisted in making a number of lamps similar to the commercial CL-2 style and introducing given amounts of gas impurity before the lamps were sealed off exhaust. They were then burned on life test at normal current for time intervals of 300 to 500 hours during which

TABLE II.

Current	Voltage	Watts in near infra-red	Conversion efficiency	Equivalent tungsten watts
4.0 amp.	15.0 volts	12.5 watts	21.0%	400 watts
5.0	15.5	16.3	21.0	525
5.5	16.5	19.6	21.6	630
6.0	17.0	21.2	20.8	675
6.5	17.5	24.1	21.2	755

[2] F. L. Mohler, Bur. Stand. J. Research **9**, 25 and 493 (1932).
[3] M. Beutell, Ann. d. Physik **36**, 533 (1939).
[4] F. L. Mohler, J. Opt. Soc. Am. **29**, 152 (1939).

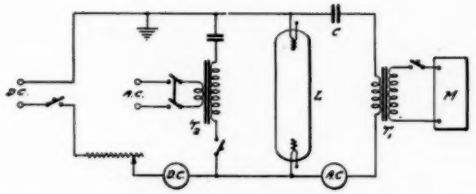


FIG. 2. Wiring diagram for operating cesium lamp. L: cesium lamp, M: modulator, C: modulation condenser of 200 to 500 mf, T_1: modulation transformer, T_2: starting transformer.

time output characteristics were periodically measured.

It was found that oxygen, even in minute amounts, is deleterious and reduces the infra-red output because of its strong quenching action on the resonance radiation of ionized cesium vapor. It should be reduced to the lowest practical limit which can be effected with excellent lamp manufacturing technique aided by the use of a suitable getter as described below.

Small amounts of nitrogen—up to 0.03 percent—gave results equally as good as pure argon gas up to 500 hours life. Lamps with 0.3 percent nitrogen in the argon gas gave only about one-third the normal output.

Carbon dioxide gas caused a reduced infra-red output probably because it dissociated into carbon and free oxygen gas.

Hydrogen gas has a beneficial effect in low concentrations as additions to argon gas. Lamps filled with a mixture of 0.006 percent hydrogen in pure argon showed an increase in the infra-red output of 30 to 50 percent throughout the several hundred hours burning on life test. All lamps burned with approximately the same voltage as the control lamps, hence the efficiency is greatly improved.

It is desirable to have a suitable getter in these lamps to remove the last traces of oxygen in the sealed off lamps. Zirconium or tantalum metal forming part of the electrode structure may increase the infra-red output an appreciable amount—of the order of 10 to 20 percent. These getters, like hydrogen, remove or inactivate small amounts of oxygen from the cesium arc.

STARTING 100-WATT CESIUM VAPOR LAMPS

Cesium vapor lamps have many peculiar characteristics, one of which is the method of starting the lamps. The wiring diagram of Fig. 2 shows the essential connections. The filaments are preheated with separate filament lighting transformers at 6 amperes 2.5 volts for one minute after which a starting current of one ampere from a 300-volt transformer produces an a.c. arc that is allowed to burn for about one minute. Then the d.c. supply voltage is applied to the lamp and the current regulated to its normal value with the control resistance. The high voltage a.c. starting circuit and filament heating circuits are disconnected and the lamp allowed to burn about 15 minutes to attain thermal equilibrium before a.c. modulation is applied.

MODULATION EXPERIMENTS

The infra-red output of our experimental lamps measured in our laboratory with a thermopile and suitable filters or with a cesium photocell showed that the required total infra-red energy was satisfactory. Permission was received to have Dr. Phillips Thomas at the Westinghouse Research Laboratories in East Pittsburgh make measurements on the frequency response throughout the audio-range. He had suitable equipment that was already set up, and in a few hours on August 4, 1944 measurements were made on three lamps using 800-cycle key signals, voice, phonograph record transmissions, and audio-frequencies of 200 to 10,000 cycles per second from an oscillator. The sound transmitted by the invisible infra-red radiations was of excellent quality. The frequency response throughout the audible range showed that high quality transmission with good efficiency could be expected up to 100 percent current modulation. Distortion appears above 100 percent current modulation due to harmonics, but the lamp will not be extinguished with overmodulation.

The circuit shown in Fig. 3 can be used to

TABLE III.

Frequency	Modulation Westinghouse	Modulation Michigan
200	75	90
500	73	80
1000	72	70
2000	65	60
3000	58	35
5000	45	25

measure both the d.c. and a.c. component of light emitted by the lamp while it is being modulated. The lamps were burned in a Dewar flask with 100 percent current modulation, and all gave comparable results. The accompanying Table III gives data showing the modulation factor for selected frequencies as determined at the Westinghouse Laboratories and also at the University of Michigan 10 days later by Dr. J. G. Black by an entirely different set of apparatus and a different method. Modulation factor is taken as the percentage light modulation for 100 percent current modulation.

These data show that high modulation factors are obtained even with our early experimental lamps. An ordinary 60-watt tungsten filament lamp can be burned on d.c. and readily modulated with audiofrequencies to 6000 cycles per second, but the modulation efficiency is very low because of heat capacity and thermal inertia of the tungsten filament. At 1000 cycles the cesium vapor lamp is about 1000 times as efficient as the tungsten filament lamp.

Two lamps of the 100-watt size can be burned in series or in parallel and modulated with the same power unit. While series operation has little to recommend it since longer lamps of higher wattage can be made quite easily, parallel operation will provide for greater reliability and insurance against failure without any mechanical difficulties of synchronization.

500-WATT LAMPS

Research on the 100-watt cesium vapor lamps was later expanded to include 500-watt and 50-watt sizes. The lamps are essentially similar in structure to the 100-watt size but have their own peculiarities. The 500-watt size was made for use without a reflector and must be burned in a vertical position. The arc stream is 2 inches in diameter, 17 inches long, and requires 10 amperes at 50 volts. Because of its length the lamp is more efficient than the 100-watt size as the data in Table IV indicate. The lamp is fitted with a base at each end of the lamp to avoid arc displacement to the side of bulb because of the magnetic effect of the current passing through a conductor near the inner envelope. This problem was not important in the smaller lamps but becomes serious in lamps carrying 10 to 20 amperes arc current. Lamps greater than 3 inches in diameter tend to become unstable even with special electrodes that were designed to keep the arc stable.

TABLE IV. Characteristics of 500-watt cesium vapor lamp.

Current (amperes)	10.0	9.0	8.0
Voltage (volts)	49.5	47.5	46.5
Total infra-red watts	172	149	121
Efficiency	34.8%	35.0%	35.3%
"Equivalent" watts	5800	4750	4150

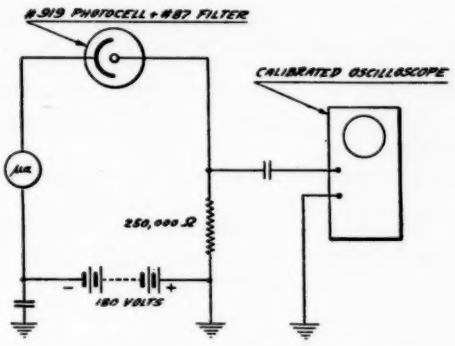


FIG. 3. Circuit used to measure d.c. and a.c. components of infra-red radiation from cesium vapor lamps.

The nominal rating of this size lamp is 10 amperes 50 volts d.c. with 100 percent a.c. modulation for code signal transmission. With modulation current superimposed upon the steady d.c., the lamp voltage is 7 to 8 volts less as measured on a d.c. meter. There is about a 20 percent increase in wattage with 100 percent current modulation which increases the visible light and long wave infra-red, since the near infra-red as measured both with a thermopile and cesium photo-cell remains constant. The visible light increases from about 500 lumens when the lamp is operating with 9.5 amperes d.c. to 750 lumens when the lamp is operating with 100 percent 800-cycle modulation current superimposed upon the d.c. current. This represents the difference in light output between a 40-watt and a 60-watt household incandescent lamp. The greatest part of the increase in energy is converted to long wave-length infra-red that raises the bulb temperature 15° to 20°C.

A "standby" current of 5 amperes may be used to keep the lamp warm, but it cannot be operated at full output for about 10 minutes

TABLE V.

Current (amp.)	4.0	3.5
Voltage (volts)	12.5	12.0
Infra-red watts	9.3	7.6
Efficiency (percent of input)	18.6	18.2
"Equivalent" watts	308	254

after normal power is supplied to the lamp. At 5 amperes the infra-red output is about 30 percent of the normal equilibrium value but 80 percent of rated output is attained in three to four minutes.

Even though a security filter may not be necessary, it is highly desirable to surround these lamps with a transparent glass cylinder to form a protecting shield from cold weather and high winds since the radiation from the outer jacket helps to maintain high cesium vapor pressure in the lamp. This condition is not encountered in the smaller lamps that are used within a reflector housing.

50-WATT LAMPS

A small size 50-watt lamp was developed for the Army Air Corps. This lamp had an arc length of only two inches with an over-all length of 6.5 inches including the base pins. The inner bulb was one inch in diameter and the outer bulb 1.5 inches in diameter. The lamp can be modulated with a.c. throughout the audio-range, and can be burned in any position. These lamps can be conveniently burned on a 24-volt d.c. supply line, but the starting cycle is a little longer than when used on a 120-volt d.c. supply. A lamp was burned on life test with 3.5 amperes d.c., 2.45 amperes 1500-cycle a.c. modulation for over 200 hours with constant output. With 4.0 amperes d.c. and 2.8 amperes 1500-cycle a.c., lamps have burned over 100 hours with approximately constant output. In these lamps a slight change in bulb shape to approximate an isothermal surface will increase the infra-red output and efficiency an appreciable amount. Table V gives data on the characteristics of the 50-watt lamps.

Immediately after the tests at the University of Michigan on August 15, 1944 several lamps were sent to Northwestern University where for some time a group of men under the direction of Dr. W. S. Huxford had been developing a communication system under an NDRC contract using a different type of lamp that did not meet all of the Navy's requirements. Our experimental lamps of 5.5 amperes current rating matched their power supply and searchlight reflector, and in a few days an operable field unit demonstrated that this system formed a practical means of communication. Two-way ship to shore voice and code communication by invisible infra-red light beam was demonstrated on the Atlantic Ocean, October 18 and 19, 1944 with equipment built at Northwestern University.

The advantage of using an invisible light beam as a carrier wave of modulated a.c. currents is that such a system can be made into a secret system that cannot be detected or jammed like radio. It is limited to horizon distances and the light can be used like a searchlight beam if desired. These lamps may be adapted to ship to shore use in crowded harbors where radio might be objectionable, and in convoy work.

ACKNOWLEDGMENTS

This project extended from April, 1944 to October, 1945 and required the assistance of men in our own organization as well as those in other laboratories mentioned above. The problem was suggested to us by Mr. Joseph Ballam at the Bureau of Ships who coordinated lamp development at the Westinghouse Lamp Division and equipment development at Northwestern University. Experimental work involved in conducting field tests at the Navy's test station and at the Naval Research Laboratory was done largely by Mr. E. Goldstein.

After field tests demonstrated the feasibility of using our 100-watt cesium vapor lamps in a secret communication system, an order was taken to make them on a commercial scale. Adapting the experimental lamps to factory production methods was assigned to Mr. T. C. Retzer, who had to make a new product rapidly, on existing equipment, in reasonable quantity and with excellent quality. Engineering development on the 500-watt size lamp occupied the full time of Mr. C. M. Rively throughout the cesium lamp project. Much of the work involved in making experimental lamps was done by Mr. A. J. Oszy.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 36, NUMBER 10 OCTOBER, 1946

A Multi-Step Sector Photometer

J. A. VAN DEN AKKER
Physics Department, The Institute of Paper Chemistry, Appleton, Wisconsin
(Received June 14, 1946)

DURING the past several years the application of our spectrograph[1] has been considerably extended through the use of a sector photometer having the following uses and characteristics: (a) determination of absorption spectra; (b) determination of spectral reflectance; (c) comparison of light sources (to obtain relative spectral energy distribution); (d) production on the photographic plate, with a single exposure, of eleven extinction levels ($E=0$, 0.1, 0.2, . . . 1.0), with intervening strips of the sample spectrum; (e) high photographic speed; and (f) simplicity of design and absence of complex and costly optical components.

A number of photometers designed for use in conjunction with spectrographs are described in the literature. The principles of operation of several of the better known photometers are presented in such well-known works as those prepared by Twyman and Allsopp[2] and Brode.[3] Harrison[4] has classified and critically discussed the many methods developed for photographic photometry.

The method employed in most existing photometers is that of the rapidly rotating sector. In designing a new photometer based on this principle, the main objective was the elimination or minimization of certain rather obvious shortcomings of the usual arrangement. One of these is complexity of optical systems designed to divide the radiation from the light source (for passage through matched absorption cells) and then to bring the beams together in close juxtaposition at the slit of the spectrograph. Such systems are costly, are sometimes difficult to place in proper alignment, and may absorb or scatter a substantial fraction of the radiation. Photometers utilizing optical systems of this kind are not readily adapted to photometric work other than determination of spectral absorption. Perhaps the most serious objection to be raised against the ordinary sector photometer is slowness of procedure; only one extinction level is produced for each exposure and sector setting. Distinct advances in speed of procedure were made by Twyman, Spencer, and Harvey[5] and by O'Brien,[6] who devised ingenious optical systems enabling one to obtain a series of extinction levels in a single exposure. The design of the new photometer incorporates the feature of multiplicity of extinction levels, but does so without requiring the fabrication of complex absorption cells or prisms.

METHOD

As in other sector photometers, radiant energy from the sample[7] is compared with definite fractions of the energy from the standard[7] by means of a rapidly rotating sector disk having known angular openings, and through the technique of locating "match points" on the spectrum plate.[4] The distinguishing feature of the method is the provision, through rapid mechanical oscillation (of frequency the same as the revolutionary speed of the sector disk), for alternately placing the sample and standard on the optic axis of the spectrograph. Depending on the particular use of the photometer, the alternate placement of sample and standard on the optic axis is *actual* or *virtual*. In the former case, the sample and standard are oscillated or rotated upon the axis (depending, again, on the use of the photometer) and, in the latter case, the radia-

[1] Medium quartz spectrograph, manufactured by the Bausch and Lomb Optical Company, Rochester, New York.

[2] F. Twyman and C. B. Allsopp, *The Practice of Absorption Spectrophotometry* (Adam Hilger, Ltd., London, 1934).

[3] Wallace R. Brode, *Chemical Spectroscopy* (John Wiley and Sons, Inc., New York, 1934).

[4] George R. Harrison, "Current advances in photographic photometry," J. Opt. Soc. Am. **24**, 59–72 (1934).

[5] F. Twyman, L. J. Spencer, and A. Harvey, "Rapid spectrophotometry with bi-multiple spectra and a new type of wedge cell," Trans. Opt. Soc. (London) **33**, 37–53 (1931).

[6] B. O'Brien, "A null method of photographic spectrophotometry with a spiral aperture sector disk," J. Opt. Soc. Am. **22**, 426A (1932).

[7] *Sample* is here used in a general sense; it may refer to a transparent solid or liquid, a solid reflecting body, or to a light source. *Standard* refers to the corresponding reference body—that is, air, or matched cell containing the solvent, a standard of reflectance, or a standard light source.

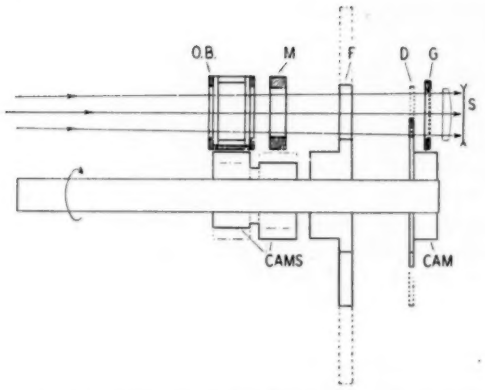


FIG. 1. Schematic side view of the photometer, showing the relationship of principal parts.

tions from fixed sample and standard are alternately deflected down the axis by means of an oscillating pair of fused silica deviating prisms. This use of an oscillating pair of prisms is similar to that of the oscillating quartz biprism in the recording spectrophotometer developed by Harrison and Bentley.[8]

In the apparatus which forms the subject of the present paper, when the sample is on the axis a uniform opening in the sector disk exposes all points in the slit of the spectrograph to the rays from the sample for the same period of time. However, a masking grate (to be described later in the article) shades the slit in eleven equally spaced segments, and permits the rays to pass through the slit in the ten intervening segments. During the next half-cycle of events, the standard is located on the axis, and the masking grate, which oscillates in phase with the sample and standard, now permits rays to pass through those segments of the slit which were previously shaded, and blocks rays directed toward those segments which were previously exposed. The half of the sector disk which is now in play has a graded series of angular openings. The largest of these openings (corresponding to zero extinction) is equal to the uniform opening in the other half of the disk (100° in the present instrument), and the remaining openings are successively smaller, to give uniform steps in extinction (or, if desired, in energy). There is a one-to-one correspondence between the exposed segments of the slit and the graded series of openings; radiation controlled by the nth angular opening falls upon the nth exposed segment, and the radial limits of the nth angular opening center on the shaded segments flanking the nth exposed segment. There are eleven angular openings and, correspondingly, eleven exposed segments of the slit.

When employed for the determination of spectral absorption or reflectance, a single light source is used and, hence, variability of the source produces no observable error. The errors, difficulties, and energy losses associated with beam splitting are absent in most uses of the method. Densities in the photographic emulsion are compared at very nearly the same point on the plate, and the rotational speed of the sector (about 1700 r.p.m.) is sufficient for reasonably good adherence to the reciprocity law.[9]

DESIGN OF THE PHOTOMETER

The design and construction of a photometer requiring rapid oscillation of mechanical parts—and, in particular, rapid oscillation of absorption cells—was approached with some hesitancy. On first consideration, it appeared that the machinery would be awkward and troublesome, and it was feared that a small bubble of air or vapor in the top of an absorption cell might produce an extensive turbulent zone filled with tiny bubbles. A very simple, feasible design for the mechanical drive of the oscillating parts suggested itself, however, and it has developed that a small bubble in the top of an absorption cell causes no difficulty. With regard to the latter, hundreds of absorption spectra (including a number which involved the use of such volatile solvents as hexane and ethyl ether) have now been determined, and no difficulty of any sort has been encountered which could be attributed to oscillation of the absorption cells.

The disposition of the principal parts of the photometer are shown in a schematic sketch presented in Fig. 1. For the purposes of illustration, it will be assumed that the instrument is

[8] George R. Harrison and E. P. Bentley, "An improved high speed recording photometer," J. Opt. Soc. Am. **30**, 290–294 (1940).

[9] J. H. Webb, "The relationship between reciprocity law failure and the intermittency effect in photographic exposure," J. Opt. Soc. Am. **23**, 157–169 (1933).

arranged for the determination of absorption spectra. Rays from the source (ordinarily an iron arc in our work) first pass through the oscillating box, *O.B.*, and the double cell which it holds, then through the rectangular opening in a compensating mass, *M*, which oscillates 180° out of phase with the oscillating box, then through the large angular openings in a sectored "flywheel," *F*, which serves to ensure very nearly constant speed of the photometer during one revolution (the openings in the flywheel are sufficiently large that rays arriving at the slit cannot be intercepted), then through the sector disk, *D*, the masking grate, *G*, and the quartz lens and slit of the spectrograph. The direction of oscillation is perpendicular to the plane of the drawing. The shaft of the photometer is driven by belt connection to an induction motor.

The oscillations are produced by eccentric cams mounted on the photometer shaft, as shown in Fig. 2. The drive for the compensating mass is similar to that shown in (a) for the oscillating box; the center of its cam is displaced 180° from that of the cam which drives the oscillating box. The amplitude of the motions of the oscillating box and the compensating mass is only 0.11 inch; the motion is approximately sinusoidal, and the frequency is approximately 28 c.p.s. The compensating mass is a rectangular brass frame of mass similar to that of the oscillating box and its typical contents. Its purpose is, of course, to minimize vibration in the base of the instrument.

The drive for the masking grate, *G*, shown in (b) of Fig. 2, is similar to that for the oscillating box. The motion of the masking grate is in phase with that of the oscillating box, and its amplitude is very small, being only 0.027 inch. Because of the small amplitude and mass of this system, no compensation has been provided. The photometer shaft system is statically balanced. When in operation, the vibration produced in the spectrograph is entirely negligible.

The frame of the masking grate is supported and oscillated through three steel points, *P*, at the upper ends of the steel reeds, as shown in (b) of Fig. 2. The reeds are under small flexural stress, thus providing a light clamping force on the frame. This method of mounting produces a purely translational oscillation of the masking grate, which is essential to the production on the photographic plate of sharply separated strips of the sample and standard spectra. When viewed from the side during oscillation, the masking grate appears to be at rest—there is no perceptible vertical component of the motion. Another desirable feature of this method of mounting is that the grate can be removed easily and replaced for cleaning with a camel's-hair brush.

In order that the rotational speed of the photometer would be uniform throughout one revolution, the flexural rigidities of the steel reeds shown in Fig. 2 were so chosen as to result in natural vibrational frequencies of the three oscillated parts close to the driven frequency. It was realized that an off-resonance condition could easily obtain and, hence, provision was made for the shaft to carry the sectored flywheel to which reference has already been made. In calculating the moment of inertia of the flywheel, a frequency equal to only one half the normal or resonant frequency was taken, and it was assumed that the flywheel should be sufficiently large to hold the speed variation during one cycle to within 0.5 percent. At or near resonant frequency, speed variation during a revolution should then be entirely negligible. The angular openings in the flywheel are each 110° and are symmetrically disposed with respect to the sector disk.

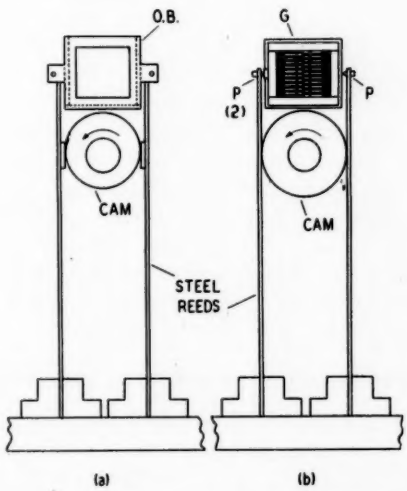


FIG. 2. On-axis views of the drives for the oscillating box and the masking grate.

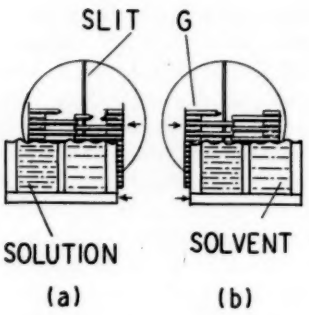


FIG. 3. Disposition of the absorption cells, masking grate, and the spectrograph slit at the extrema of the oscillatory motion.

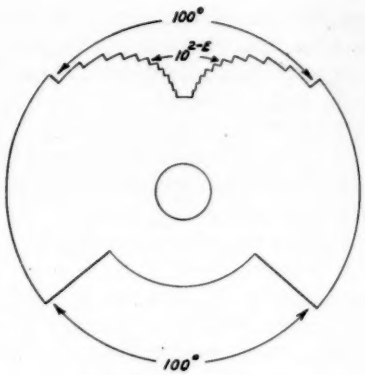


FIG. 4. The sector disk.

The relationships between the positions of the absorption cells, masking grate, and spectrograph slit at the extrema of the oscillatory motions may be seen by reference to Fig. 3, (a) and (b). When in position (a) the rays are cut by the graded series of angular openings in the sector disk (shown in Fig. 4), and when in position (b) all rays are cut by the uniform, 100-degree opening of the sector disk.

A photograph of the photometer mounted on the spectrograph is given in Fig. 5. The base of the instrument is only 6 by 6 inches. Locating pins enable an operator quickly to mount the photometer on the spectrograph in proper alignment. The belt shown in Fig. 5 proved to be troublesome and connection through the pulley on the flywheel hub has been discarded. Recently, the flexible cable attachment shown on the end of the shaft nearest the reader was removed and replaced with a special 5-inch pulley for a $\frac{3}{8}$-inch V-belt. This pulley consists only of a hub, two narrow spokes, and the sheave, thus permitting free passage of off-axis rays when the photometer is employed for determination of spectral energy distribution. The hub of the pulley is fitted with a short shaft extension to provide rotation of auxiliary equipment needed in determination of spectral reflectivity.

DETERMINATION OF SPECTRAL ABSORPTION

If it is desired to obtain the spectral transmission of a gelatin or glass filter, or other transparent body, it is often possible to cut a small rectangle of the material (0.5 by 1 inch) and insert the piece in the sample half of the oscillating box. If the substance is a solution, it is usually desirable to determine transmittance—i.e., the ratio of transmission of the solution to that of the solvent. Matched double cells are then employed. Two forms of cells are shown in the photograph reproduced in Fig. 6. One of these is a double cell and is made of polished plates of fused silica; the other pair are separate absorption cells mounted in staggered arrangement in a light brass frame. The staggered arrangement permits the left side wall of the rear cell to fall behind the right side wall of the front cell, thus

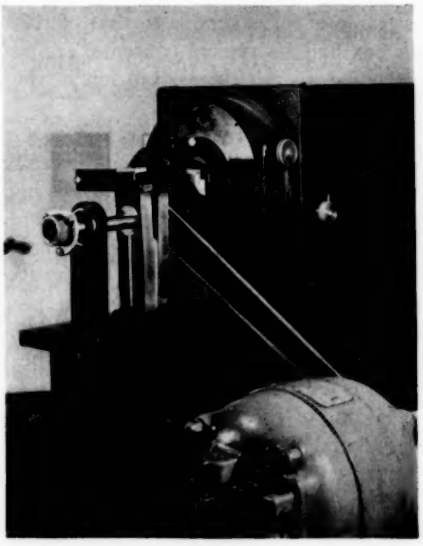


FIG. 5. The multistep sector photometer mounted on the spectrograph.

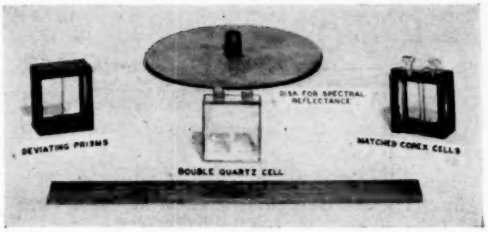


FIG. 6. Some of the auxiliary devices used with the photometer: deviating prisms, disk for spectral reflectance, and absorption cells. For comparison of size, a 6-inch rule is included in the photograph.

minimizing the separation of the cells. A double quartz cell (not shown in Fig. 6), having an inside spacing of one millimeter, is comprised of two square plates of quartz (1 by 1 inch) cemented with phenol-formaldehyde resin to an Invar frame. In addition to providing a seal around the edges, the frame provides a thin central partition. A small threaded hole in the top of each compartment permits filling, emptying, and flushing by means of hypodermic syringes. These holes are fitted with threaded plugs. This double cell has been surprisingly convenient in use. Only one solvent, cellosolve, damaged the seal, and the damage became apparent only after prolonged use of this powerful solvent; small bits of the "flash" of the polymerized resin on the inside edges of the compartments became loosened—apparently, the damage was caused by slight swelling rather than dissolving of the resin.

When it is undesirable to prepare a small piece of filter to place in the oscillating box, or if it is required to employ absorption cells of unusually great inside length, the transparent body may be placed at C in the arrangement shown schematically in Fig. 7. Light from the source S is reflected from front-faced reflectors m' and m, and enters the photometer by way of a pair of fused silica deviating prisms, P, in the oscillating box, $O.B$. These prisms serve alternately to deviate the two beams down the optic axis of the spectrograph and into the slit. The angle of each prism is 20°. As shown in Fig. 6, the prisms are mounted in a brass frame (which fits snugly in the oscillating box).

If it is desired to obtain spectral transmittance with the arrangement shown in Fig. 7, a matched cell containing the solvent is placed at C'.

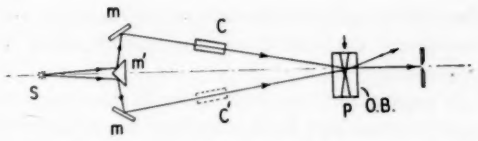


FIG. 7. Arrangement for determining spectral absorption of filters or cells that are too large for placement in the oscillating box of the photometer.

In view of the fact that fused silica absorbs ultraviolet light, particularly in the wave-length region approaching 200 mμ, the system shown in Fig. 7 must be carefully arranged to equate the path lengths of the two beams in the corresponding prisms. This is done rather quickly by turning the photometer by hand, aligning the arc, and adjusting the position of the photometer to obtain equal travel of the shadow of the line of contact of the two prisms on either side of the slit. Actually, the compensation is strictly valid (on the assumption that the absorption coefficients in the two prisms are equal) for low values of extinction, for then the weighted mean path lengths through the prisms are equal; however, at an extinction of 1.0, the path length of the standard beam is very nearly constant at $t+2A$ tan 10° [where t is the thickness of the prisms at their line of contact (0.12 cm) and A is the amplitude of the oscillation (0.28 cm)] or 0.219 cm, whereas the path length of the sample beam varies between $t+2A$ tan 10° and $t+2A$ tan 10° cos 50°—that is, between 0.219 and 0.183 cm. It is a fortunate characteristic of sinusoidal motion that displacements corresponding to a small range of angles about the peak displacement do not differ greatly from the latter. The simple time-weighted[10] mean path length of the sample beam is 0.207 cm. Ordinarily, the correction is small, and is not made; as a numerical illustration, let us take a wave-length at which the transmission of 0.219 cm of the fused silica is 0.900. Calculation shows that the transmission of 0.207 cm of the fused silica would be 0.905. This indicates a correction of about 0.5 percent, which is well within the limits of accuracy of photographic photometry. The correction should, of course, be applied to higher extinction values

[10] The strict treatment would include the exponential term, exp $(-\mu x)$, in the weighting function, where μ is the absorption coefficient for the radiation in the fused silica, and x is the instantaneous path length in the fused silica.

when working in a wave-length region where the absorption coefficient of the fused silica is appreciable.

In exploratory work of a qualitative nature, good results have been obtained by dispensing with the reflecting surfaces m' and orienting the surfaces at m to reflect the radiation from the source directly through C and C' to the pair of deviating prisms. The two beams then leave the source in widely divergent directions, and one can no longer assume that the intensities of the beams are equal, in view of asymmetry of the radiation from (most) arcs.

Virtually all of our work on spectral absorption is done with the normal arrangement shown in Fig. 1—that is, with the absorption cells or filter placed in the oscillating box. Radiation from the arc passes directly down the optic axis through the photometer to the spectrograph. Using an iron arc at a current of 3.5 amperes (d.c.), gap setting of about 2 mm, and an arc to slit distance of about 65 cm, exposure times vary typically from about 5 seconds to 2 minutes (in accurate work, short exposure times should be avoided). The most appropriate exposure time depends on the photographic emulsion, the wave-length range of primary interest, and the nature of the solvent and its thickness.

DETERMINATION OF SPECTRAL REFLECTANCE

The accuracy of photographic photometry is not sufficiently high to satisfy the requirements of careful color analysis. Frequently, however, data on the spectral reflectivity of solid bodies in the ultraviolet, visible, or near infra-red regions are required for technical or analytical purposes. Ordinarily, the accuracy of a sector photometer is more than ample for such purposes.

The arrangement employed for the determination of spectral reflectance is shown schematically in Fig. 8. The sample and standard surfaces are mounted on a metal disk, which is driven by a short drive shaft connected to the photometer shaft through flexible couplings, J. The standard half of the disk is smoked with magnesium oxide. The method of mounting the sample on the other half of the disk depends upon the nature of the material. If it is an opaque sheet, a half-disk is cut from the sheet and glued to the metal disk. In other cases, the material can be painted or

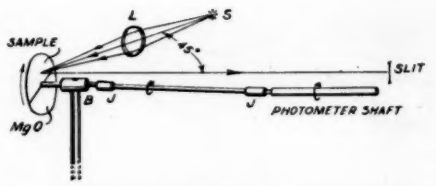


FIG. 8. Schematic drawing of the arrangement for determining spectral reflectance.

coated upon the metal disk. Small areas can be employed when it is desired to minimize centrifugal force (which has never proved to be troublesome) by the simple expedient of raising the axis of rotation of the metal disk almost to the point at which the top of the sleeve bearing, B, shown in Fig. 8 intercepts the reflected beam to the slit.

Light from the source S is condensed by means of a quartz lens, L, to illuminate a small area[11] on the surfaces of the sample and standard at the point where the optic axis intersects the disk. The angle of incidence may be varied, but is ordinarily set at 45°.

The photometer is now employed with the oscillating box empty; the rotation of the disk alternates the placement on the optic axis of the sample and standard in synchronism with the sector disk and the masking grate. In view of the fact that the angular extent of the sample and standard is 180°, the adjustment of the metal disk is not critical, and a lag due to frictional drag of the order of 20° or more would be permissible. Stroboscopic examination shows, however, that lag is negligible.

In one study, which involved low reflectance in the ultraviolet region, the number of match point levels was increased by cutting the incident beam with a rotating screen of low transmission. The screen cut the beam only when the standard was on the optic axis. (It might be appropriate to mention at this point that a screen has proved useful in studies of low transmission of transparent sheets—the disk shown in Fig. 8 was replaced with a simple frame shaped like a capital *theta*, which served to carry the screen; the screen covered only the half of the frame

[11] A fortunate result of the rotation of the sample and standard surfaces is that the energy of the condensed radiation is distributed around an annulus of appreciable circumference.

rotating in phase with the graded half of the sector disk.)

When determining spectral reflectance, typical exposure times vary from a few minutes to about $\frac{1}{4}$ hour.

Comparison of Light Sources

The photometer may be employed to compare the spectral energy distribution of radiation from a "sample" source with that of the radiation from a "standard" source. The pair of deviating prisms shown in Figs. 6 and 7 is used alternately to place the virtual images of the two sources on the optic axis. With reference to Fig. 7, the light sources are located in the positions occupied by the reflectors *m* at equal, or known, distances from the slit (as measured along the deviated rays).

When the relative spectral energy distributions of the two sources differ very greatly, a suitable filter of known spectral transmission is placed in either the sample or standard beam. This practice reduces the number of exposures required to obtain the complete energy distribution curve for the sample source.

If the source is extensive in area, its radiation must be integrated or condensed to emerge from a small area, or the source must be masked with an opaque screen having an opening of a few millimeters in diameter.

One of our most interesting applications of the photometer was in connection with a study of fluorescent map papers. The sample source consisted of a 5-millimeter square of the paper mounted on the end of a rod, illuminated by a nearby Mazda BH-4 lamp. The reference source was a piece of flashed diffusing glass covered with black paper having a 5-millimeter square hole, illuminated from the rear with a Mazda 1061 lamp (concentrated tungsten filament) enclosed in a lamp house to reduce error caused by scattered light. The distance between the tungsten filament lamp and the diffusing glass was adjusted in such manner that the maximum in the energy distribution curve for the fluorescent radiation was slightly less than the corresponding ordinate of the curve for the reference source. The distances between the sources and the slit of the spectrograph were the same—about 35 cm. The distance between the axis of the BH-4 lamp and the specimen of map paper was 6.2 cm. Under these conditions, a not quite satisfactory exposure was obtained in 30 minutes, and satisfactory density was obtained after an exposure of one hour. Sampling error was minimized by replacing the specimen with a fresh specimen at 15-minute intervals.

There is a good chance that the application of the photometer to the comparison of light sources can be extended to emission spectrum analysis. The two sources would be similar arcs containing the same amount of specimen material, except that the material in the standard arc would contain a known added amount of the element under investigation. We have not yet been able to schedule work on the development of an arc-pair in which very nearly the same conditions of excitation would obtain. However, other aspects of the method appear to be satisfactory. We are currently using the internal standard method with a logarithmic sector and high voltage a.c. arc.

PERFORMANCE OF THE PHOTOMETER

The accuracy of data obtained with the sector photometer is about that to be expected with eye-match photometry with photographic plates of good contrast. In the visible region, where the spectral transmission of a filter determined with the photometer can be compared with that obtained on the General Electric recording spectrophotometer, rather good agreement is obtained—provided that the transmission curve is free of sharp curvatures, which introduce some error in the recorded curve when, as in the case of our recording spectrophotometer, the slit widths are set for a range of 10 mμ. Most points determined with the photometer fall within a range of approximately one percent (transmission unit) or less about the recorded curve, and occasional deviations of about two percent are noted. In view of the lower contrast of photographic emulsions at shorter wave-lengths, appreciably lower accuracy is to be expected far in the ultraviolet region, especially when the wave-length is of the order of 200 mμ. The accuracy in determining a spectral absorption curve can, of course, be increased by making two or three exposures with different cell thicknesses or concentrations calculated to yield data that

are fairly evenly distributed with respect to extinction coefficient.

The mechanical operation of the photometer has been exceptionally free of difficulties. As previously stated, hundreds of absorption spectra have been determined, and no mechanical failures have developed.

REMARKS ON THE READING OF THE SPECTROGRAMS

As mentioned earlier in this article, the spectrogram consists of eleven strips of the "standard" spectra alternating in close juxtaposition with ten strips of the "sample" spectrum. The plate is placed emulsion side down on an illuminating box, and the match points are located with a pen on the glass side of the plate facing the observer. After reading and recording the wave-lengths corresponding to the match points, the inked points are wiped off with a damp cloth if it is desired to make two or more determinations of the match points.

In the interest of objectivity, the match points are located with a light-colored but opaque mask which exposes only three strips—*viz.*, a standard strip and the immediately adjacent sample strips—and the inked points are placed over the middle or standard strip. The mask hides nearby match points. In addition, the mask eliminates an illusion that is readily apparent when the spectrogram is viewed in its entirety. Although the sample strips are all of very nearly the same density, the presence of the intervening standard strips of graded density causes neighboring sample strips to have the appearance of different density. The illusion vanishes when the mask is employed. Better accuracy and comfort obtain when the match point determinations are made with the mask illuminated by a nearby desk lamp arranged to equalize the luminance of the viewed strips and the surface of the mask (as opposed to making the observations in a dark room).

A magnifying lens is employed when the extinction curve is very steep. However, where the slope of the curve is low, we believe that match points can be determined more accurately with the unaided eye. The observer then scans the strips in a slow, oscillatory manner, permitting the center of attention to vary between points to the left and right of the match point. Very probably the better discrimination results from the use of both eyes and the more extensive field of view.

Wave-length is read by laying a transparent scale upon the photographic plate. The lines on the scale have curvature closely similar to that of the spectrum lines, and are spaced at one-millimicron intervals in the range from 200 to 360 mμ. A central line on the scale is placed over the 0.5 extinction strip, thus making it easy for the observer to read the extinction as well as the wave-length. The scale is properly located by placing the division at 310 mμ over the conspicuous iron lines at that wave-length.

ACKNOWLEDGMENTS

The care exercised by Mr. Carleton H. Root in the construction and preliminary testing of the sector photometer is gratefully acknowledged. Dr. Ralph E. Glading, who first employed the photometer in an extensive study, prepared the wave-length scale, the use of which has greatly facilitated our work.

Plastic Filters for the Visible and Near Infra-Red Regions*†

JOHN H. SHENK, EDWIN S. HODGE, ROBERT J. MORRIS, EDWARD E. PICKETT, AND WALLACE R. BRODE
Department of Chemistry, The Ohio State University, Columbus, Ohio

INTRODUCTION

IN connection with certain devices which were developed for use during the war, sources of near infra-red radiation were required. Since all common infra-red emitting sources also produce visible light, the most convenient and satisfactory method for obtaining near infra-red radiation is to eliminate visible light by placing a filter over an incandescent lamp, carbon arc, or other suitable light source. Filters for this purpose may be produced by incorporating dyes in plastic materials, subliming phthalocyanines and similar materials upon glass, coloring glass, and by the Christiansen filter effect.

Difficulties involved in the production of efficient filters indicated that colored glass and dyed plastics were the most desirable. Although the preparation and use of glass has been fairly well explored by glass manufacturers, glass filters with steep absorption curves are not available for the near infra-red region. This report is concerned with the search for dyes and plastics which can be combined to give filters with the necessary transmission characteristics, satisfactory stability to heat and intense light, and resistance to weathering.

A study of several hundred dyestuffs revealed that a relatively low percentage of available dyes is suitable for the preparation of filters. Certain azo dyes combine the necessary transmission properties with reasonable heat stability. The fact that they can easily be converted into an alcohol soluble form makes them especially valuable since they can be incorporated into a number of liquid plastics for preparing waterproof filters. The presence of solvents in the liquid plastics precludes the possibility of casting. Therefore, the plastic dye solution is coated upon a suitable glass support by spraying or pouring. The coating is allowed to dry in the air for a few hours prior to baking.

Particular attention has been given to the preparation and testing of filters for the near infra-red region. These filters were to transmit a maximum of infra-red with a minimum transmission of visual radiation. It has been shown that efficiency of such filters for a particular wave-length cut-off depends upon the shape of the absorption curve, which in turn depends upon the molecular structure of the dye used.

DYE STUDIES

Dyes must possess certain characteristics for the successful preparation of colored plastic, light absorbing filters. Among these characteristics are compatibility of the dyes with suitable plastics, ability to withstand dry heat at 120°C for long periods, stability to strong tungsten and arc radiations, the sharpness with which they divide the spectrum, the spectral region covered, and the availability of these substances in relatively large amounts at a reasonable cost.

The cooperation of dye manufacturers made it possible to investigate a large number of commercial dyes. With the exception of Naphthol Green B, it was found that certain azo dyes made the more satisfactory light filters for the near infra-red region.

The commercial dyestuffs were qualitatively checked for shading agents and those showing little or no shading were selected for use. The structures were assumed to be as listed in the *Colour Index*.[1] The dyes were converted into diorthotolylguanidine (DOTG) salts by the method of Rose.[2] The dyes were made spirit soluble by this process, which also served to eliminate all but traces of inorganic salts. Absorption spectra of these dyes in organic solvents were obtained by means of a Beckman quartz spectrophotometer.

* The work described in this paper was carried out, in whole or in part, under Contract OEMsr-987 between the Office of Scientific Research and Development and the Ohio State University Research Foundation. The project is being continued under a U. S. Army Engineer Board contract with the Ohio State University Research Foundation.

† Paper presented at the Winter Meeting of the Optical Society of America, Cleveland, Ohio, March 7–9, 1946.

[1] F. M. Rowe, Society of Dyers and Colourists, *Colour Index* (Chorley and Pickersgill, Ltd., Leeds, 1924).

[2] R. E. Rose, Ind. Eng. Chem. **25**, 1028 (1933) U. S. Patent 1,674,128 (1928).

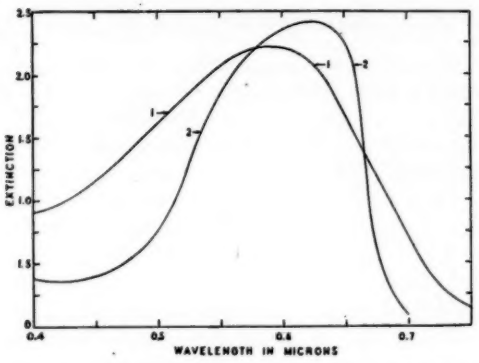


FIG. 1. A comparison of the absorption bands of a suitable (1) and an unsuitable (2) dye for an infra-red filter.

(1) Pontamine Diazo Blue NA (C.I. 317)
(2) Pontamine Sky Blue FF (C.I. 518)

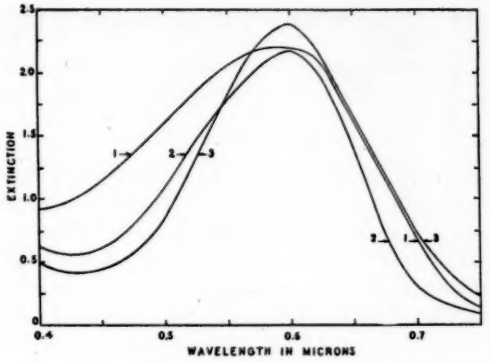


FIG. 2. The effect of structural variation on the slope of the long wave-length side of the absorption band. Concentration 0.0002 *M*.

(1) Pontamine Diazo Blue NA (C.I. 317)
(2) Sulphon Cyanine New GA (C.I. 288)
(3) Chlorantine Fast Blue 4 GL (C.I. 533)

A small number of these dyes combined a suitably sharp cut-off with the necessary spirit solubility. Certain structural types of azo colors appear to be more effective for the preparation of light filters. Dyes suitable for infra-red filters showed a gradual sloping of the absorption band on the long wave-length side. This absorption characteristic was far more important than the position of the absorption maximum, since the maxima recorded for about 30 dyes varied by only 50 millimicrons and yet only a few of these dyes were suitable for use in filters for the infra-red region. For example: Pontamine Diazo Blue NA (C.I. 317)[3] with an absorption maximum at 570 mμ, will give a filter cut-off in concentrated solution at a considerably longer wave-length than Pontamine Sky Blue FF (C.I. 518) with an absorption maximum at 620 mμ but with an extremely sharp band on the long wave-length side (see Fig. 1). Since the absorption maxima of most of the available azo dyestuffs occur at values lower than 0.6 micron, it is necessary to use relatively high concentrations in order to obtain filter cut-offs beyond 0.7 micron (Figs. 2 and 3).

The most satisfactory diazo dyestuffs were usually connected through a naphthalene nucleus in the 1,4 positions (Table I). This structural peculiarity appears to enhance the broadening of the absorption band and reduces the slope of the band on the long wave-length side. This effect increases as the exterior radicals are progressively changed from benzene to naphthalene derivatives. This same effect is also shown in a similar trisazo dye (C.I. 533) in which the three

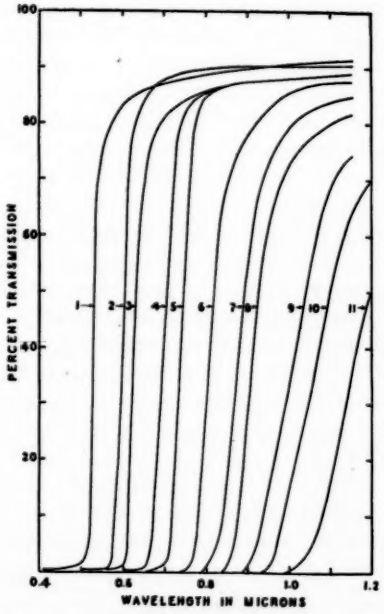
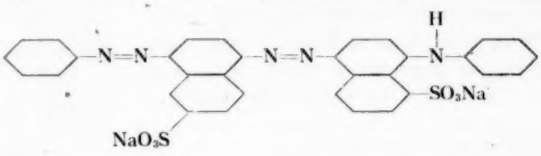
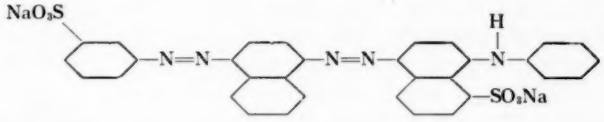
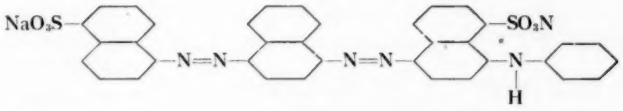
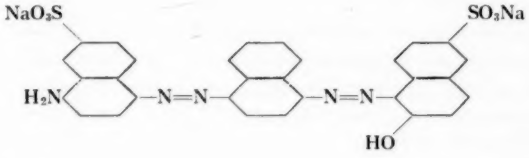
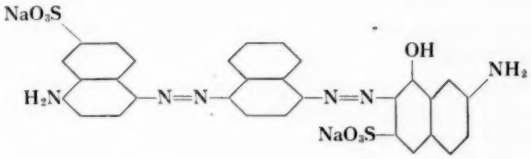
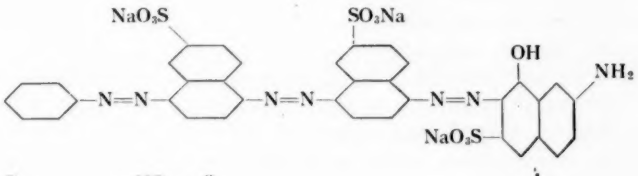
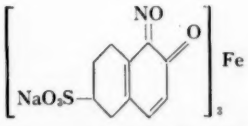


FIG. 3. Filter transmission curves for the visible and near infra-red regions.

The dyes used are as follows: (1) Luxol Fast Yellow T, (2 and 3) Luxol Fast Red B, (4) Luxol Fast Blue AR, (5) Luxol Brilliant Green BL, (6) Chlorantine Fast Blue 4 GL (DOTG Salt), (7 and 8) Pontamine Diazo Blue NA (DOTG salt), (9, 10, and 11) Pontamine Diazo Blue NA (DOTG Salt) and Luxol Fast Green B, in a 4 to 1 mixture, respectively.

[3] C.I. indicates the color index number as given by Rowe.[1]

TABLE I. The structural formulas of the most satisfactory dyes for near infra-red filters.

Name	Color index No.	Structural formulas
Sulphon Cyanine New GA	288	
Cloth Fast Blue R	289	
Pontacyl Fast Black BBN	307	
Calcomine Diazo Blue B2RD	316	
Pontamine Diazo Blue NA	317	
Chlorantine Fast Blue 4GL	533	
Naphthol Green B	5	

azo groups are coupled 1,4 to two naphthalene nuclei (Table I and Fig. 2).

From a careful examination of the readily available spirit soluble dyes, the Du Pont series of Luxol dyes was found to be quite satisfactory as additions to the list of converted azo dyestuffs (Table II).

SELECTION OF A SUITABLE PLASTIC BASE

The plastic solvent for the dyestuff must be transparent in the spectral region used and must be capable of hardening to form a smooth surface and a tenacious bond to a suitably transparent support. In addition, it must be sufficiently resistant to weathering, thermal shock

strong light radiations, the abrasion of ordinary use, and dry heat up to 120°C.

A modified urea formaldehyde plastic, Beetle *230–8*,[4] and a modified melamine formaldehyde plastic, Melmac *599–8*,[4] were found to be suitable plastic bases. As these when used alone or in combination did not bond sufficiently well to a glass support to meet all requirements, they were further modified by the addition of 30 to 70 percent of an alkyd resin, *Rezyl 330–5*.[4] The alkyd resin plasticizes both the urea and the melamine types of coating so that they harden at 100°C to a tough, slightly resilient film. Without the use of Rezyl, the Melmac and Beetle plastics set on baking to an extremely hard brittle solid.

TABLE II. Most satisfactory dyes for filters in the visible and infra-red regions.

Name of dye	Color index No.	Cut-off* range	Dye ratings
Luxol Fast Yellow T		0.50 –0.525 mu	Excellent
Luxol Fast Orange GS		0.525–0.575	Good
Luxol Fast Red B		0.550–0.625	Excellent
Luxol Fast Blue AR		0.625–0.675	Good
Luxol Brilliant Green BL		0.675–0.70	Excellent
Luxol Fast Black L		0.680–0.725	Good
Luxol Fast Blue G		0.70 –0.76	Excellent
Luxol Fast Green B		0.90 –1.15	Good
Sulphon Cyanine GA New	288	0.725–0.775	Good
Cloth Fast Black	289	0.725–0.80	Good
Chlorantine Fast Blue 4 GL	533	0.750–0.80	Excellent
Pontacyl Fast Black BBN	307	0.75 –0.825	Good
Calcomine Blue B2RD	316	0.775–0.850	Excellent
Pontamine Diazo Blue NA	317	0.775–0.905	Excellent
Naphthol Green B	5	0.90 –1.15	Good

* Cut-off—The term as used here means the wave-length at which the spectral transmission of the completed solid filter is 1 percent.

PROCEDURE FOR THE PREPARATION OF INFRA-RED FILTERS

A suitable transparent support was necessary as a base for the plastic filters. Glass was satisfactory for nearly all filter applications; however, in certain cases, methyl methacrylate sheets were found to be a better support. Ordinary green plate glass should not be used because of a transmission loss of approximately 20 percent in the infra-red region.

The method used in the manufacture of the filters is easily adaptable to either a laboratory or a limited commercial scale. The *Melmac 599–8* was mixed with 30 to 70 percent of *Rezyl 330–5*. One percent of *Guai-A-Phene*[5] was added as an anti-skinning agent to insure a smooth surface free from prohibitive patterning. Beckamine[6] accelerator (2 to 3 percent) can be added if a lower curing temperature is required. Such temperatures are often desirable to prepare more pliable surface coatings. This also allows the use of clear plastic supports that would deform at temperatures above 80°C.

The dye-plastic dope is prepared by introduction of the finely divided dye as a dry powder into the viscous plastic medium with continuous mechanical stirring. The viscosity of the mixture can be suitably reduced with butanol or xylene if the plastic becomes too viscous to be conveniently stirred by ordinary methods. Stirring is continued for 6 to 12 hours during which time further addition of solvent can be made to maintain a convenient fluidity. The colored dope is then centrifuged in an ordinary laboratory centrifuge or, for large quantities, a Sharples continuous supercentrifuge is used. This procedure eliminates undissolved dye and any salt suspended in the dope, thus avoiding scattering of light by the finished filter.

Suitable amounts of colored plastic, properly regulated in viscosity and concentration, are dispensed in the center of glass plates or disks and allowed to flow to the edge. The plates or disks are placed on previously leveled trays and are allowed to air dry for 6 to 12 hours. They are then placed on carefully leveled oven trays for periods of 6 to 12 hours at 100°C. The period of drying and baking increases with the size of the filter and the amount of material required. A three-point adjustable screw support is satisfactory for keeping the trays in a level position for drying and baking. Glass plates, upon which 75 to 100 small coated filters can be placed, may be handled as a unit.

The reasons for rejection of filters are usually unevenness of film thickness and pinholes. The evenness of coat is dependent upon the careful leveling of the filters during the periods of air drying and baking. Pinholes can be avoided by the elimination of bubbles on the disk. Acetone or other surface active solvents have proved most successful in exploding bubbles that occur in the dope before pouring or in the freshly poured film. When thick coats are required to meet certain filter specifications, cellulose tape

[4] Manufactured by American Cyanamid Company.
[5] Manufactured by Southern Pine Chemical Company.
[6] Manufactured by the Reichhold Chemical Company.

around the edge of the disk successfully retains the plastic on the surface of the support until its viscosity has increased sufficiently by air drying to permit the removal of the tape whereupon the edge coating can then be completed. Edge coating of the filters is accomplished by pulling the semi-set plastic coat down over the edge of the disk with a brush after the film has air dried for a period of from 12 to 24 hours.

The preparation of colored liquid plastic dope for use in spraying can be done by diluting the original dope 20 to 30 percent by a thinner such as butyl Cellosolve. An ordinary spray gun can then be used in applying thin coatings to irregular surfaces. For best results the object to be sprayed should be revolved in a suitable fashion while the coat or coats are sprayed on. Each spray coat is cured by placing the article in an oven maintained at 100°C for 20 minutes to an hour before a second spray coat is applied.

Another type of filter can be prepared by the use of 2 percent of *Beckamine Accelerator* in the plastic dope which is then poured on methyl methacrylate sheets similar to the method of preparing the filters on glass. These films if allowed to stand for one hour in the air followed by a 2 to 4 hour baking period at 70°C can be stripped off the methyl methacrylate support very easily. These flexible stripped films are extremely resistant to wrinkling and crushing.

ABSORPTION SPECTRA OF PREPARED FILTERS

Filters with satisfactory cut-off values from 0.4 to 1 micron were prepared by varying the concentration of the original stock solutions. Transmission curves of prepared filters covering the visible and near infra-red regions are shown in Fig. 3. The filters were measured against air in a Beckman quartz spectrophotometer. To improve the accuracy of the instrument between 0.9 and the limit of 1.15 microns, a Corning Glass No. 254, was used in the filter slide.

In a consideration of the efficiency of light filters for use in the infra-red, it is as important to absorb the visible portion of the spectrum as it is to permit a high transmittancy of the desired infra-red wave-lengths. The efficiency of this process is indicated by the sharpness of the toe of the spectral transmission curves recorded for the various infra-red filters.

The effectiveness of the prepared filters for the visible and near infra-red portions of the spectrum is shown in Fig. 3. A comparison of the effectiveness of these filters with other types of prepared filters is illustrated in Fig. 4. The sharpness of the toe and the total transmission should be noted.

EVALUATION OF INFRA-RED FILTERS

Since infra-red filters are prepared to transmit a maximum of infra-red radiation and a minimum of visual radiation for a particular system, a detailed knowledge of the source, receiver, and eye characteristics must be known in order to properly evaluate a filter. The infra-red filters described above were designed to be used with a tungsten filament lamp as a source and a cesium oxide photo-cell or phosphor receiver. In Fig. 5, are shown the spectral sensitivity of an *S*-2 surface cesium oxide photo-cell, the spectral sensitivity of the human eye as determined by Goodeve,[7] the spectral distribution of the radiation energy from a tungsten lamp of color temperature 2848° as given by Forsythe,[8] and a transmission curve of one of the filters described above. The values in each case are relative. Although the eye curve seems to merge with the base line, Griffin, Hubbard, and Wald[9] have shown that the eye can detect light out as far as one micron. The largest proportion of the visible light is transmitted at the toe of the absorption curve and, therefore, dyes with sharp cut-offs are sought. It should be noted that although steepness of the absorption curve is desired, if the visual transmission increases faster than the eye's threshold decreases, little is gained by increasing the steepness.

The effectiveness of an infra-red filter for the near infra-red region can be described by comparing the transmission of radiation affecting the eye with the transmission of radiation which will actuate the receiver. A quantitative evaluation of the effective visual transmission of a filter, T_v, can be determined from its spectral transmission curve for light of any spectral intensity distribution by means of the following

[7] C. F. Goodeve, Proc. Roy. Soc. **A155**, 644 (1936).
[8] W. E. Forsythe, *Measurement of Radiant Energy* (McGraw-Hill Book Company, Inc., New York, 1937).
[9] Griffin, Hubbard, and Wald, Optical Society of America Meeting, Cleveland, March, 1946.

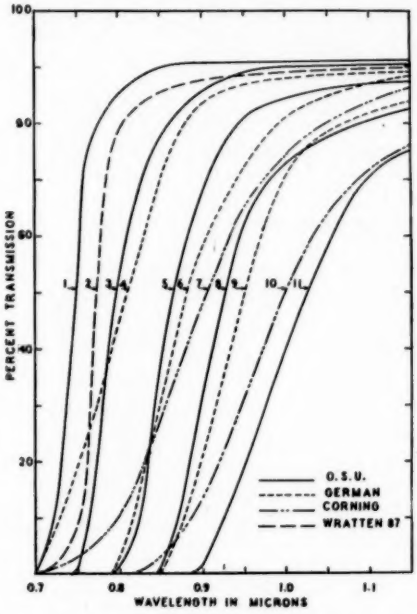


FIG. 4. A comparison of infra-red filters. (1, 3, 5, 8, and 11) Ohio State University filters, (2) Wratten 87 filter, (4, 6, and 9) German filters, (7) Corning 2550 glass filter, (10) Corning 2540 glass filter.

equation:[10]

$$T_v = \frac{\int \psi_\lambda T_\lambda I_\lambda d\lambda}{\int \psi_\lambda I_\lambda d\lambda},$$

where for any wave-length λ, ψ_λ is the relative luminosity function for the eye, T_λ is the transmission of the filter, and I_λ is the spectral intensity of the source (see Fig. 5). The effective infra-red transmission, T_r, may be determined by a similar equation in which the eye response is replaced by the spectral responsivity of the receiver, R_λ.

$$T_r = \frac{\int R_\lambda T_\lambda I_\lambda d\lambda}{\int R_\lambda I_\lambda d\lambda}.$$

Theoretically the integrals are to be taken from

[10] R. L. Tousey and D. E. Mitchell, Optical Society of America Meeting, Cleveland, March, 1946.

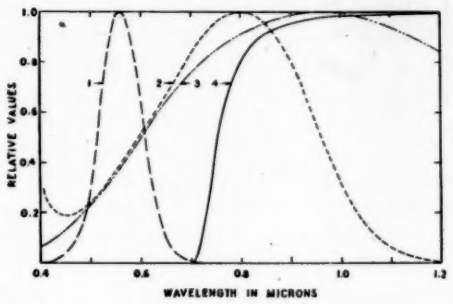


FIG. 5. (1) Spectral sensitivity of the human eye, (2) spectral sensitivity of the *S*-2 cesium oxide photo-cell, (3) radiant energy distribution for tungsten at 2848°, and (4) filter transmission curve.

$\lambda=0$ to $\lambda=\infty$, but the integration need only extend over the limited wave-length range in which the increments make a significant contribution. Values of T_λ, I_λ, R_λ, and ψ_λ may be obtained from Figs. 3 and 5. It will be noted that T_r differs for different filters when used with different receivers. For a particular receiver, one can conveniently express the merit, M, of an infra-red filter by use of the ratio T_r/T_v. The merit ratings increase as the transmission curve is moved to longer wave-lengths

TABLE III. Transmission and merit values of infra-red filters.

Concentration	T_r	T_v	M
	Luxol Fast Black L		
40	0.565	4.0×10^{-5}	1.4×10^{4}
60	0.54	2.3×10^{-5}	2.4×10^{4}
80	0.505	1.3×10^{-5}	3.9×10^{4}
100	0.47	7.4×10^{-5}	6.3×10^{4}
	Luxol Fast Blue G		
40	0.525	6.8×10^{-6}	7.7×10^{4}
60	0.49	3.5×10^{-6}	1.4×10^{5}
80	0.465	2.5×10^{-6}	1.9×10^{5}
100	0.415	1.3×10^{-6}	3.2×10^{5}
	Pontamine Diazo Blue NA (DOTG Salt)		
10	0.48	1.3×10^{-5}	3.7×10^{4}
20	0.35	6.8×10^{-7}	5.15×10^{5}
40	0.25	6.7×10^{-8}	3.7×10^{6}
60	0.195	1.4×10^{-8}	1.4×10^{7}
80	0.17	7.0×10^{-9}	2.4×10^{7}
100	0.14	3.0×10^{-9}	4.7×10^{7}
	Values obtained by varying film thickness		
100	0.133	2.3×10^{-9}	5.8×10^{7}
100	0.120	1.4×10^{-9}	8.6×10^{7}
100	0.103	7.1×10^{-10}	1.6×10^{8}
100	0.095	3.8×10^{-10}	2.5×10^{8}
100	0.084	2.4×10^{-10}	3.5×10^{8}

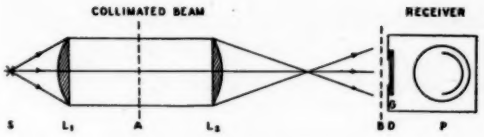


FIG. 6. Infra-red photometer.

since the visible transmission is decreased at a greater rate than the infra-red transmission as shown in Table III.

In practice, it is more convenient to measure T_v and T_r than to carry out the necessary integration. The visual transmission values in Table III were obtained with a comparator constructed on the principle of the McBeth illuminometer and which had been calibrated with a set of standards furnished by the Naval Research Laboratory. The beam from a ribbon filament lamp passes through the filter and a lens, which focuses the image of the filament on one-half of a two field Lummer-Brodhun cube. For comparison, the other half is illuminated from the side by a 15-candlepower auto lamp on a sliding scale with an opal glass to provide uniform illumination. One of the several dark red Wratten filters, depending upon the density required, is placed in the latter light path. A variable sector in the direct test beam serves to allow measurements of very light filters.

Figure 6 shows schematically the arrangement of the apparatus for determining the effective infra-red transmission, T_r. Light from a tungsten source, S, at a color temperature of 2848°K is brought to bear upon the ground glass, G, immediately in front of the photo-cell, P, which is connected to a simple DC amplifier circuit. The transmission, T_r, is expressed as the ratio of the reading with the filter in position B to the reading with the filter removed from the beam.

This instrument also provides a simple method for quantitatively measuring the amount to light scattering by inserting the filter into the collimated beam at position A. The percentage

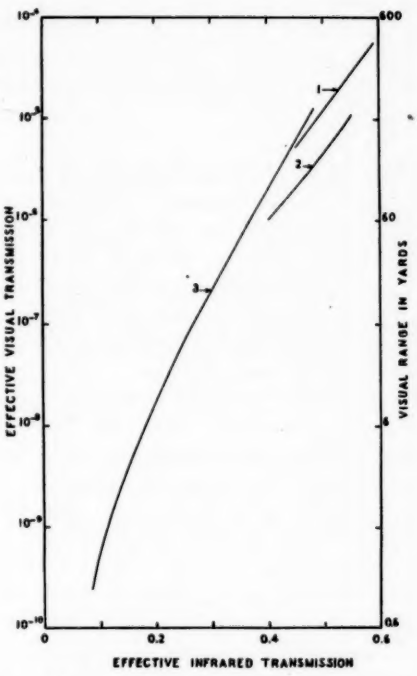


FIG. 7. Filter evaluation curves for (1) Luxol Fast Black L, (2) Luxol Fast Blue G, (3) Pontamine Diazo Blue NA (DOTG salt).

of scattering is determined from the difference in the readings with the filter at positions A and B.

Values of T_v, T_r, and M, are given in Table III for infra-red filters prepared from several concentrations of the most satisfactory dyes. The concentrations designated 100 percent are essentially saturated solutions. The other concentrations are expressed as percentages of the saturated solutions.

The values in the above table are plotted in Fig. 7. Visual ranges, when used with a 100-candlepower source, have been calculated and are shown opposite the effective visual transmission values.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 36, NUMBER 10 OCTOBER, 1946

Radioactive Light Sources

M. BLAU AND I. FEUER
Research Laboratories, International Rare Metals Refinery, Inc., New York, New York
(Received August 12, 1946)

THE scope of the present paper is to emphasize the importance and usefulness of radioactive preparations together with adequate luminescent screens as constant light sources. In a paper on the measurement of polonium with the aid of the multiplier photo-cell[1] the feasibility of such radioactive light standards was discussed.[2]

It is obvious that the fluorescent effect of radioactive radiations, especially that of the highly ionizing alpha-radiation, can be used as a light source. For a long time luminous compound mixed with radioactive material has served as a paint for dials of watches and instruments. However, the light emitted by these radioactive luminous compounds decreases with time, even if radium (half-life 1590 years) is used, because the continuous bombardment of alpha-particles produces changes in the crystal structure and in the optical properties of the luminous compound. More detailed investigations are now being undertaken in order to find the exact nature of these phenomena.

On the other hand, if the radioactive source is separated from the luminous compound and irradiation of the screen only takes place during relatively short intervals, the light output is extremely constant. Certain phenomena of fatigue which occur after a somewhat longer or stronger irradiation are transitory, and the initial efficiency is restored after a short rest.

The distance between radioactive source and screen, the fluorescent material, the type and intensity of the radioactive source, and the geometric arrangement, depend upon the use of this radioactive light source.

The most convenient radioactive preparation —if absolute constancy is required—is radium plated on a metal foil in such a manner that there is no escape of emanation and a minimum absorption for the emitted alpha-particles.[3] The constants of these radium foils (number and energy of the alpha-particles emitted) can be tested very accurately.

The only inconvenience of these preparations is the penetrating gamma-radiation which may be disturbing in case of radioactive light standards of great intensity and luminous surface, making necessary the use of stronger radium preparations.

Instead of radium, polonium can be used, which practically emits no penetrating radiation. Because of the limited half-life of polonium (140 days) these light sources are not constant, but the respective light intensity at the given time can be calculated exactly if the initial intensity is known, because the light effect is proportional to the number of alpha-particles hitting the screen. For that reason the intensity of the light source L_t at the time t (in days) is given by

$$L_t = L_0 \exp\left(-\frac{0.693}{140}\cdot t\right) \qquad (1)$$

where L_0 is the initial light intensity.

The simplest radioactive light source consists of a radioactive preparation and a screen arranged at a certain distance in a light tight box; between the radioactive source and screen is a metal shutter which protects the screen from radiation when the apparatus is not in use.

In order to have light sources of various degrees of luminosity the arrangement shown in Fig. 1 or Fig. 2 may be used.

In Fig. 1 the distance between preparation and screen can be varied or the velocity of the alpha-particles can be decreased by inserting absorption foils, so that the range of the alpha-particles within the screen is diminished.

In Fig. 2[4] the distance between radioactive source and screen is kept constant and the

[1] M. Blau and B. Dreyfus, Rev. Sci. Inst. **16**, 245 (1945).
[2] In a private communication, Mr. C. W. Walhausen, Chief Chemist of U. S. Radium, informed us that he had, for some time, been using radioactive light standards as a source for light.
[3] Compare reference 2.
[4] A similar arrangement was first used for investigation on the homogeneity of the alpha-radiation of various emitters by B. Karlik, Wien. Akad. Ber. **142**, 115 (1933.)

velocity of the alpha-particles can be varied by changing the pressure within the tube.

The luminosity of the screen hit by alpha-particles depends on the screen material, on the number of alpha-particles, and on their velocity or range within the screen.

Especially for zinc sulfide screens, the light output is very high; about 20–80 percent of the alpha-particles absorbed in the material are transformed into light depending upon the conditions.[5] The energy of alpha-particles incident upon the screen after having spanned a distance A in air of standard atmospheric pressure is given by

$$E_{R-a}=K^{\frac{1}{2}}\frac{m(R-a)^{\frac{1}{2}}}{2}.$$

In this equation m is the mass of the alpha-particles and K a constant; the relation between the velocity of the alpha-particles and their range R is given by $v^3=KR$.

The light effect produced by the alpha-particles is proportional to the loss of energy sustained in the luminous material. The energy loss is given by

$$\frac{dE_{R-a}}{dx}=\frac{K^{\frac{1}{2}}m}{3}(R-a)^{-\frac{1}{2}}$$

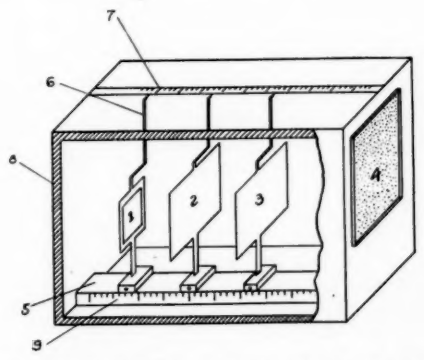


FIG. 1. Optical light standard unit (regular type). 1. Radioactive source (radium or polonium) which is mounted on a stage placed on optical bench (5). 2. Aluminum absorbing screen of known absorbing power. 3. A steel screen or preferably a shutter which may be controlled from the outside (similar to type in cameras). 4. Fluorescent screen. (This entire unit may be mounted in any position in order to accomplish desired function: e.g., light standards, control, or regulating medium.) 5. Optical bench. 6. Extension of stage mounted in order to read external scale. 7. Upper external scale. 8. End of casing. 9. Lower scale of optical bench.

[5] N. Riehl, Ann. d. Physik **29**, 636 (1937).

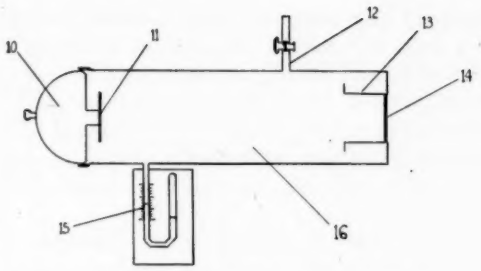


FIG. 2. Radioactive optical light standard (pressure controlled type). 10. Cover of glass or metal tube. 11. Radioactive source (mounted at constant distance from screen 14). 12. Stopcock to control gas pressure. 13. Canalization (metal) for the alpha-particles. 14. Fluorescent screen (ZnS type). 15. Manometer for determining pressure. 16. Gas pressure or vacuum.

and the total light effect produced by alpha-particles with the rest range $R-a$ is proportional to the integral of this expression. Therefore

$$L=c\int_0^{R-a}dE_{R-a}=c(R-a)^{\frac{1}{2}}. \qquad (2)$$

Equation (2) is only a first approximation as the "range" of the alpha-particles is not completely independent of the velocity and besides depends upon the nature of the absorbing material; especially for alpha-particles of small velocity which are at the end of their range, Eq. (2) can no longer be applied. It has to be substituted by a more complicated expression taking into account the exact theory of the passage of alpha-particles through matter, or by figures found experimentally.

In Fig. 3 curves II and I show the decrease of light intensity with the distance between source and screen. The ordinates are light intensities in arbitrary units. The measurements in II were made with a polonium preparation and the one in I with a radium preparation. The light effects due to beta- and gamma-radiation in the latter case were measured separately and subtracted from the measured values.

Curves 2 and 1 of Fig. 4 represent measurements with polonium and radium sources having a constant distance from the screen; the rest range of the alpha-particles in the screen was decreased using absorption foils of aluminum of known stopping power. In the case of 1 the effect due to beta- and gamma-radiation was again subtracted. The difference between the

polonium and radium curves is caused by the fact that we have in the case of polonium a practically uniform alpha-radiation of the range of 3.87 cm (at atmospheric pressure and 15°C) while in the case of radium we have to deal with the alpha-particles of radium with 3.39 cm, radon with 4.12 cm, RaA with 4.72 and RaC with 6.96 cm.

In Fig. 5 the logarithms of the ordinates of curve 2 of Fig. 4 are plotted against the rest range $(R-a)$. The theoretical points are calculated assuming that within the luminous compound alpha-particles of all ranges between 0 and $R-a$ are uniformly distributed, as the path of the alpha-particles in the air (before reaching the screen) depends from the angle of incidence. In this case we can suppose that the light effect will be proportional to

$$c\int_0^{R-a}(r-a)^{\frac{1}{2}}d(r-a)=R^{3/2}+c \qquad (3)$$

In the first part of the curve the agreement between experimental and theoretical values is fair, but with shorter velocities the deviation of the theoretical curve becomes more and more pronounced due to the reasons mentioned above.

In the experiments a special zinc sulphide screen which emits a blue violet light under the bombardment of alpha-particles was used.

In order to calibrate a radioactive light standard in units of brightness, it is necessary to compare it with a calibrated light source of the same color. If this is done once, this source can then serve as a secondary standard for the calibration of others with the same spectroscopic properties but of different intensities or different geometrical arrangement.

The brightness of radioactive light sources of different colors can be compared only if the spectral sensitivity of the photo-cell (or whatever light measuring device is used) is exactly known; otherwise, it is necessary to calibrate each color separately with the respective standardized light source.

The brightness of the radioactive light source depends upon the luminous material, the surface density of the radioactive source and its position relative to the screen.

The factor which determines the transformation of alpha-particle energy in light is very different for the various luminous compounds. Moreover, the transmission factor for the light emitted in the fluorescent process also plays an important role in the final output of light. The optimum light output for the individual luminous material is reached if the thickness of the screen is equivalent to the maximum range of the alpha-particles. Because, it is the stopping power which determines the maximum range

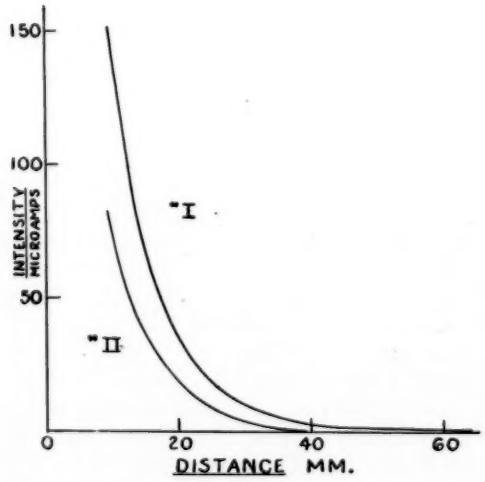


FIG. 3. Alpha-intensity of radioactive sources *vs.* distance. Curve I: radium source, β- and γ-radiation corrected for. Curve II: polonium source. Units: light source measured in arbitrary microamp., distance measured in millimeters.

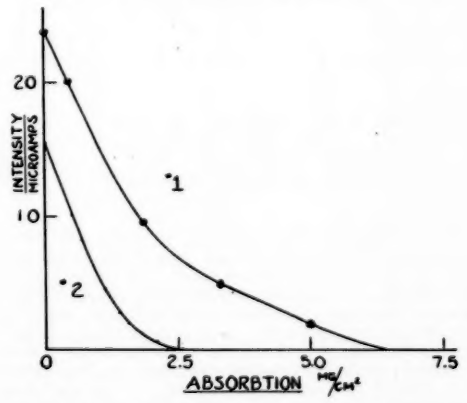


FIG. 4. Light intensity of radioactive substances *vs.* absorption by aluminum foil. Curve 1: radium source. Curve 2: polonium source. Units: light-arbitrary microamp.; absorption—mg/cm².

(and is different for the various luminous compounds), this optimum thickness varies from one screen material to the other.

Usually the best light output is reached with copper activated zinc sulphide screens (green fluorescent light), but there exist some zinc sulphide compounds which, when excited by alpha-particles, emit blue or violet light of the same or even higher brightness.

In a first approximation we can state that for the copper activated zinc sulphide screens we get a brightness of about 1 to 2 microlambert with sources of about 1 microgram radium per square cm and a distance of about 4 mm between source and screen. If polonium is used instead of radium the surface density must be about 6 microcurie per square cm in order to obtain the same brightness of 1 to 2 microlambert.

With higher surface densities of the radioactive preparation the brightness increases proportionally within a wide range until saturation is reached. The saturation point depends upon the luminous compound employed. With the blue and violet luminous compounds mentioned above, a brightness of about 300 microlambert can be reached with a polonium preparation of 1.8 millicurie per square cm showing that this great brightness is still below the saturation point. In the case of radium we would need a surface density of about 3 mg per square cm, in order to reach the same brightness. The preparation of radium foils of such high surface density is not impossible but encounters complications of a rather difficult nature.

The radioactive light standards could be very useful for colorimetric measurements and similar purposes. They have the advantage of being easier to handle than ordinary light standard as they do not involve electric currents which must be kept constant.

Besides, in the case of radioactive light standards, the light output can be varied by varying the intensity of the radioactive preparation or by absorption of the radiation, whereas in the case of light standards an increase in current or absorption influences the spectral distribution of the source.

However, the advantage of such radioactive light standards is even greater for testing work

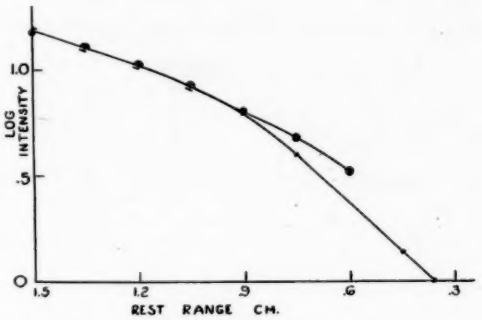


FIG. 5. Log intensity *vs.* rest range of a polonium source. 1: experimental curve ×; 2: theoretical curve ⊗; 3: where experimental and theoretical points coincide ⊗.

on cathode-ray screens or of the luminous compound used for such screens.

The spectral distribution of the fluorescent light produced by alpha-particles is very similar or identical to the one produced by cathodic rays. The luminosity obtained with cathode-rays, of a certain specification, on a luminous compound can be standardized in units of light obtained with the same luminous compound, under the bombardment of alpha-particles. For that reason this method may also be very important for the standardization of the color of luminous compounds for television purposes. The right mixture of various luminous compounds, in order to give the desired color, can be investigated in small samples outside the cathode-ray tube which in turn means a saving of time and material.

Moreover, the radioactive light standard may be useful for the judgment of the persistence of phosphors, as this persistence period is of the same order of magnitude as in the case of cathode rays.

Another advantage is that effective screen thickness of the alpha-device and cathode-ray screen can be made identical: the penetration depth for alpha-particles of RaC corresponds to that of cathodic rays of 40,000 volts, for the alpha-particles of polonium the range in zinc sulphide is about 28 microns corresponding to cathode rays of about 30,000 volts. For the comparison of cathode rays of smaller voltage we would have to use absorption foils or greater distances between source and screen in order to

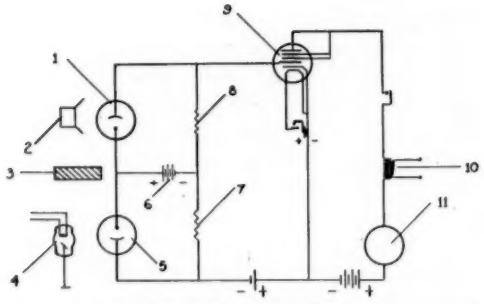


FIG. 6. Radioactive light standard as an x-ray control device. 1. Photo-tube to measure light from radioactive source. 2. Radioactive light source. 3. Lead partitions to prevent x-rays from reaching photo-tube 1. 4. X-ray apparatus. 5. Photo-tube to measure radiation from 4. 6. Batteries and 7–8 Resistances (to parallel branches of the circuit as shown). 9. Amplifier tube (No. 38 may be used). 10. Relay to control desired function. 11. Meter to observe deviations if relay 10 is omitted.

slow down the velocity of the alpha-particles hitting the screen.

A radioactive source introduced directly into the cathode-ray tube may be used in order to control and measure the thickness of the screen; as this can be done before the tube is closed and evacuated it might save time and work.

It is self-evident that instead of radium or polonium any artificial radioactive beta-emitting source can be used; however, because of the smaller fluorescent effect of these particles (about 100 times less) much stronger sources are required.

Finally we want to mention that the combination of an alpha source together with a screen might be employed for many devices where this constant source can control the variation in the sensitivity of photo-cells or other light measuring devices. As an example we give the schematic Fig. 6 where the radioactive standard is used to control and regulate the photoelectric section of the x-ray circuit. Figure 7 shows as an example a device to control the maintenance of the level of a bar, disk, or plate.

To begin with Plate A is set parallel to B. Plate B has two holes which are covered with glass plates painted with luminous compounds. Plate A has two alpha- (or beta-) radiation preparations whose radiation falls on the screens 15 and 16. The two circuits of the photo-cell are balanced, so that no current flows through the

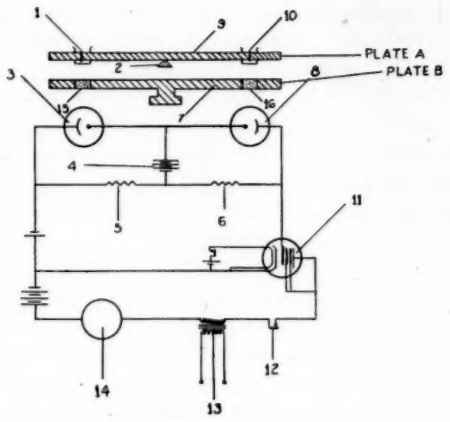


FIG. 7. Radioactive light standard as a bar level, disk, plate or weight control. Plate A—Movable plate; Plate B—Supporting plate. 1 and 10. Radioactive alpha- or beta-emitting preparations which are mounted on Plate A balanced on a fulcrum 2: These preparations irradiate screens 15 and 16, respectively. These screens are mounted on fixed support Plate B which throw light upon photo-tubes 3, and 8, respectively, which in turn regulate the relay or meter through the depicted amplifying circuit. The circuit is so regulated that when the radioactive light standards are equidistant from the screen the out is zero. When this relationship is disturbed so by tilting Plate A—a current will flow through relay 13 and meter 28. This deflection may be then corrected and controlled by suitable means or merely measured. 2. Fulcrum. 3. Photo-tube. 4. Batteries and 5–6. Resistances (to parallel branches of the circuit as shown). 7. Plate B fixed support. 8. Photo-tube. 9. Plate A movable support. 10. Radioactive light source (the tope of which may be used for weighing, etc.). 11. Amplifier tube (a No. 38 tube may be used). 12. Rheostat to control current through relay and meter. 13. Relay for control of desired function. 14. Meter to observe deflection if relay is omitted. If we employ above apparatus with meter at 0 when the bar is level, it is advisable to have a compensation circuit to observe the *direction* and magnitude of deflection. 15. Fluorescent screen (ZnS type). 16. Fluorescent screen (ZnS type).

meter. If 1 or 10 comes nearer to the screen the equilibrium is disturbed and current flows. This current can be used to work a relay which brings the Plate A back to the right position. As in this case the difference of two currents is measured we can use polonium instead of radium.

This principle can also be used as a mass production weighing control; in this case the current in the meter has to be calibrated in units of weight.

In conclusion we would like to say that radioactive light sources may be employed as a new type of light standard, control and regulating device.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 36, NUMBER 10 OCTOBER, 1946

Electrodeless Discharges at High Frequencies and Low Pressures

C. R. NISEWANGER, J. R. HOLMES, AND G. L. WEISSLER
University of Southern California, Los Angeles, California

(Received June 27, 1946)

In the last ten years electrodeless discharges have become increasingly important to the spectroscopist as light sources that possess distinct advantages over d.c. discharge tubes. An attempt was made to gather information on the characteristics of such discharges at frequencies of 10, 15, 20, and 25 megacycles and pressures ranging from 10^{-2} to 10^{-4} mm of Hg using dry air as the filling gas. It is hoped that the data presented here in form of onset and offset potentials *versus* pressure will help the spectroscopist in the design of his light sources and of the associated electrical equipment, thus taking this aspect of his research at least partly out of the realm of trial and error.—With this as our primary objective the experimental conditions were *not* appropriate for yielding data that contribute to the understanding of the fundamental mechanisms of such discharges. Therefore a detailed discussion of the shapes of the characteristic curves V_s *versus* p is omitted.

1. INTRODUCTION

IT seems worth while to review very briefly the work done by earlier investigators on glow discharge, particularly on that excited by high frequency electric fields. Such a self-sustaining discharge is usually defined as one in which the charge carriers are formed in the gas itself, in contradistinction to a non-self-sustaining discharge which is partially maintained by external sources of electrons and/or ions (generated for instance by a hot filament, by photo-electric action of external light sources, by x-rays and gamma-rays, etc.). The formation of such charge carriers is not described by a single mechanism, but rather by a series of several processes such as: impact ionization by electrons (the most important), secondary electron emission from the cathode by positive ion bombardment, and others. These processes, contributing to the formation of charge carriers, are critically dependent on the particular gas employed, on pressure, and on potentials. They are counteracted by processes leading to the loss of charge carriers such as the neutralization of electrons and ions at the electrodes, recombination of ions, and their diffusion out of the active field region.

Therefore a self-sustaining discharge can be maintained in general if the number of charge carriers formed is equal to the number lost taking all the above mentioned processes into consideration. If the number of ions and electrons formed is greater than the number lost, then a disruptive discharge in form of a spark or an arc may occur, but this is critically dependent on external circuit conditions.

These general statements apply equally well to d.c. and a.c. excitation of discharges. One of the most convenient ways of representing them is to plot a typical Paschen curve, i.e., breakdown potentials *versus* pressure (Fig. 1).

It can be shown[1] that for uniform fields (and with modifications for more general field conditions) the electrode potential necessary for a self-sustaining or even disruptive discharge is a function of the product pd, where p is the pressure and d the electrode spacing. A simple physical interpretation of Paschen's Law, $V_s = F(pd)$, is as follows. The product pd represents essentially the number of molecules encountered by a charge

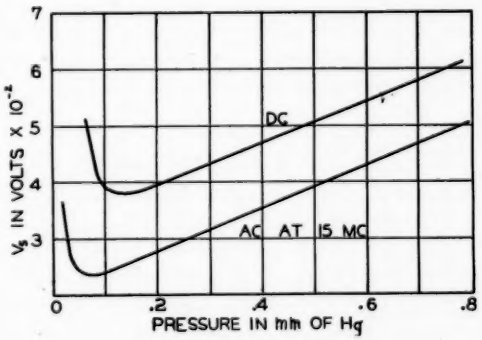


FIG. 1. Typical Pashen curves.

[1] Leonard B. Loeb, *Fundamental Processes of Electrical Discharge in Gases* (John Wiley & Sons, Inc., New York, 1939), pp. 557–8.

carrier. This in turn determines the probability of further formation of charge carriers. There will then be a specific, critical value of pd, for which the excitation potential is a minimum. At values of pd higher than this critical one, the mean free paths become shorter or the energy gained between collisions becomes smaller; therefore V_s must be increased. This accounts for the rise of the curve (Fig. 1) to the right of its minimum. Interestingly enough this portion of the curve is linear over a considerable range of values of pd, and it can be represented quite accurately by an equation

$$V_s = A + Bp,$$

where A depends critically on the gas and is a function of the frequency. The constant B is similarly dependent.

At values of pd lying to the left of the minimum, the concentration of molecules per electron mean free path decreases rapidly. Therefore V_s has to be increased to increase the ionizing probability and compensate for the decrease in collision frequency.—It is this particular region of the Paschen curve to which this investigation applies.

Discharges initiated by rapidly alternating electric fields have been investigated by Gutton,[2] Kirchner,[3] J. Thomson,[4] and others with the particular aim of casting light on certain fundamental electrical properties of gases that cannot be evaluated by the use of d.c. excitation (for instance electrode effects can be excluded with a.c. excitation). The most complete theory on such a.c. excited discharges is presented by J. Thomson,[4] but this also is considered incomplete.

At this point it may not be amiss to state that this investigation was not undertaken with the primary object of contributing data important to the understanding of the fundamental mechanisms of such discharges. As will be seen later, in the description of the discharge tubes, the closeness of the glass walls and the configuration and type of the electrodes were not appropriate for fundamental investigations.[4]

TABLE I.

Frequency mc	Diam. cm	Length cm	Turns
10	1.7	11	30
15	1.7	6	11
20	1.7	4	8
25	1.7	1.5	3

High frequency electrodeless discharges have been found empirically by spectroscopists to have certain advantages over d.c. discharges.[5–11] In any discharge tube with internal electrodes there is considerable risk of contaminating the gases under investigation because of outgassing of the electrodes. Also in cases where it is necessary to operate the discharge tube at a high temperature in order to investigate the spectra of low vapor pressure materials one is likely to have trouble with melting of internal electrodes, cracking of tungsten seals, etc.[7]

There is one line of spectroscopic research in which the use of electrodeless discharges is definitely indicated, namely investigations of the spectra of very rare materials. If small samples of an element are excited in an ordinary discharge, i.e., a Schüler tube, there is almost always such difficulty with the electrical clean-up on the metal electrodes that the sample may completely disappear in a very short time.[9] Even in an electrodeless discharge there is a small amount of clean-up of certain gases but it is very much less than in a d.c. discharge.[10] In the electrodeless discharge it is frequently possible to excite an initially smaller sample than in a d.c. discharge tube,[10] especially at very high frequencies.

It is frequently found that electrodeless discharges produce a much higher intensity of certain lines than other types of discharges.[6, 9] Especially in the case of forbidden lines has it been helpful to use a high frequency discharge.[5–8] F. A. Jenkins and S. Mrozowski have excited some forbidden lines of neutral Pb with an intensity of the order of 500 times that previously obtained.[6]

[2] Gutton, Mirta, and Ylostalo, Comptes rendus **176**, 1871 (1923); **178**, 476 (1924). C. and H. Gutton, Comptes rendus **186**, 303 (1928).

[3] F. Kirchner, Ann. d. Physik **77**, 287 (1925); **7**, 798 (1930).

[4] J. Thomson, Phil. Mag. **10**, 280 (1930); **18**, 696 (1934); **23**, 1 (1937).

[5] S. Mrozowski, "Forbidden lines in the laboratory," Rev. Mod. Phys. **16**, 153–174 (July-Oct., 1944).

[6] F. A. Jenkins and S. Mrozowski, Phys. Rev. **59**, 808 (1941).

[7] S. Mrozowski, Phys. Rev. **58**, 1086 (1940).

[8] H. Niewodniczenzki, Acta Phys. Polonica **2**, 375 (1934).

[9] J. R. Holmes, Phys. Rev. **63**, 41–46 (1943).

[10] J. H. Wiens and L. W. Alvarez, Phys. Rev. **58**, 11, 1005 (1940).

[11] S. Mrozowski, Zeits. f. Physik **112**, 223 (1939).

For the measurement of the hyperfine structure of rare isotopes the high frequency discharge has advantages as mentioned above and seems to have no serious disadvantages. Very sharp lines can be obtained if the discharge tube is immersed in liquid air. The effective temperature of an electrodeless discharge can be kept as low as that of a Schüler tube[11] and still produce high intensity lines.[9,10] Judging from the potentials we have measured across our discharge tubes, there is not likely to be much trouble with Stark broadening.

The specific objective of this investigation was to measure accurately onset and offset potentials at different frequencies for the low pressure region of the Paschen curve in order to provide data for electrodeless discharge light sources for spectroscopic investigations.

The data presented here are the onset and offset potentials for a visible glow discharge in dry air. The range of pressures used was 4.8×10^{-2} mm of Hg to 10^{-4} mm of Hg, the frequencies were 10, 15, 20, and 25 megacycles per second.

Two methods of excitation were investigated: (I) the potential was applied to ring electrodes fitted around the outside of the discharge tube, thus relying entirely on electrostatic action; and (II) the tube was surrounded by the close fitting tank coil of the oscillator and was therefore subjected to both electromagnetic and electrostatic fields.

2. DESCRIPTION OF THE APPARATUS

The apparatus employed consisted of a cylindrical glass discharge tube, a high frequency power supply, and the necessary auxiliaries for controlling and measuring the pressure, frequency, and potential.

Pressures were measured with a Pirani gauge which was calibrated against a precision McLeod gauge. Cold traps prevented the diffusion of mercury vapor into the system. Two grounded copper screen grids were inserted in the tube connecting the Pirani gauge in order to prevent ions from entering it. A grounded copper envelope shielded the gauge from the high frequency fields of the oscillator. A pressure stabilizer, consisting of a 1000-ml Pyrex glass bulb had to be added to the system in order to reduce pressure changes

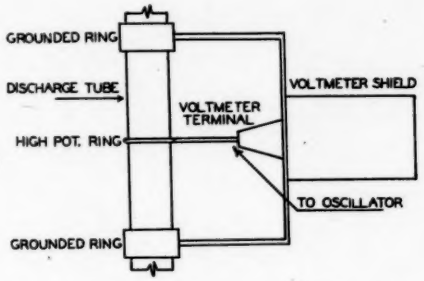


FIG. 2. External electrode system, center ring at high frequency potential, outer rings grounded.

which occurred during the operation of the discharge.

The discharge tube measured 1.2 cm inside diameter. Two different lengths, one of 20 cm and the other of 52 cm, were used. The length had no noticeable effect on the results as the tube extended well beyond the region in which the discharge occurred.

The Pirani gauge was connected to the bridge through a coaxial cable whose external conductor was grounded. By-pass condensers across the galvanometer and both ends of the coaxial cable secured the bridge against high frequency fields. A constant bridge current of 78 ma was maintained, and the readings of the slidewire (1000 divisions) settings for zero galvanometer current were taken as a measure of the pressure. After the vacuum system had been thoroughly outgassed the Pirani gauge was calibrated. During the course of the investigation the calibration was rechecked at intervals, and dial readings at any pressure did not vary more than about 2.5 divisions from the mean, the accuracy thus being within about 1.25 percent at 10^{-2} mm of Hg, 8.5 percent at 10^{-3} mm of Hg, and 25 percent at 2.5×10^{-4} mm of Hg.

Although no provision was made for temperature compensation, error due to this cause was negligible as the temperature remained between 22° and 24°C at all times.

Over 2000 volts could be obtained across the tank coil of the oscillator, and with suitable plug-in coils frequencies of 10, 15, 20, and 25 megacycles were available.

High frequency energy was transmitted to the discharge tube by clipping leads to the tank coil through mica condensers. When coil excitation

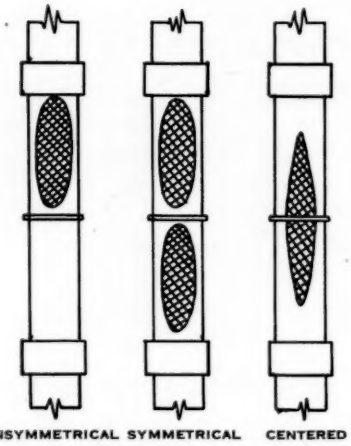


FIG. 3. Three types of glow discharge, obtained with the three-ring electrode system.

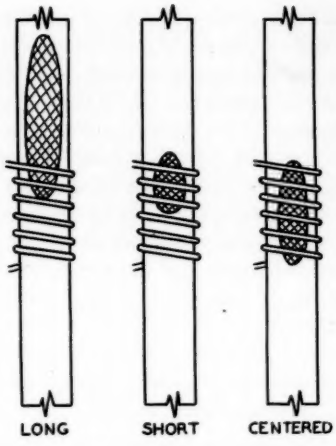


FIG. 4. Three most frequently occurring types of glow discharge, using coil excitation.

was used the entire tank circuit was removed from the chassis and transferred to a support near the vacuum system and the tank coil placed around the discharge tube. Coils used in this arrangement were wound of number 15 bare copper wire to fit closely around the discharge tube. Their dimensions are shown in Table I.

The frequency was measured with a General Radio 655A absorption type wave meter.

Several different electrode systems were tried before the three-ring system shown in Fig. 2 was adopted.

The field about any arrangement of electrodes is influenced by the presence of other conductors in the neighborhood. With a two-electrode system the discharge did not always occur between the electrodes but would sometimes occur in the stray field outside the electrodes which was not under control. (For calculation of the field about two such rings refer to J. Thomson's works.[4]) In order to control the most intense portion of the field and minimize the influence of extraneous conductors, the three-ring system (Fig. 2) was found most effective. The high frequency potential was applied to the central ring, the two outer rings being grounded; thus other conductors were relatively far away and partially shielded from the center electrode.

With the three-ring arrangement the discharge always occurred in the controlled region between the two grounded electrodes. It was not possible, however, to locate the center ring with sufficient accuracy to obtain symmetrical discharges at all times.

The vacuum tube voltmeter shield was mounted directly on the grounded electrode support and the high voltage terminal of the voltmeter supported the center ring.

A peak diode type vacuum tube voltmeter[12] was used. Since the tank coil was at high d.c. potential with respect to ground, blocking condensers were necessary to isolate the voltmeter and electrodes.

The voltmeter was calibrated on a 60-cycle source with a root mean square voltmeter. Peak voltage was assumed to be 1.414 times the r.m.s. voltage. The accuracy of the voltmeter is estimated to be within about 10 percent. Since high frequency potential measurements are usually subject to considerable error and since relative values were the important consideration in this investigation the voltmeter appeared satisfactory for the purpose.

3. EXPERIMENTAL PROCEDURE

The system was first evacuated to a high vacuum with liquid oxygen in the main trap and the McLeod gauge stopcock closed. With the pumps running, air was admitted through the dryer. This raised the pressure to somewhat

[12] John F. Rider, *Vacuum Tube Voltmeters* (John F. Rider Publisher, Inc., New York, 1940), p. 175.

above 0.1 mm and beyond the range of the investigation. Pumping continued until the pressure dropped to 4.8×10^{-2} mm, which was the upper limit of the Pirani gauge, then the main stopcock was closed, shutting the pump off from the system.

The oscillator, set at the desired frequency, was turned on and the high frequency potential across the electrodes increased, by raising the d.c. plate voltage, until the discharge started. The potential was immediately lowered until the discharge stopped. This was repeated several times with different fillings and the onset and offset potentials recorded. Thus data were obtained for four frequencies, 10, 15, 20, and 25 megacycles and for three electrode spacings, 2.2, 4.75, and 15.0 cm.

At lower pressures it was necessary to maintain a weak discharge except for very brief intervals, otherwise it could not always be restarted. This is in agreement with the findings of Kirchner[3] and Thomson[4] who suggest that a critical electron density may enter into the mechanism.

The procedure was the same when coil excitation was employed.

4. RESULTS

The onset potential was found to be quite high and erratic on the first reading but usually decreased and reached a relatively stable value after the discharge had been started and stopped a few times, provided it was restarted immediately. As expected in all cases, the onset potentials were higher than the offset potentials, diverging farther and farther as the pressure increased above 2×10^{-3} mm of Hg. At pressures below this value the onset and offset curves approach each other more and more closely, coinciding at pressures below 10^{-3} mm of Hg. This is the case for both ring and coil excitation. Figures 9 to 12 can serve as an example. In some ranges of pressure, especially at the lower frequencies, the onset potentials were high and erratic. A needle of 10 milligrams of radium placed in contact with the discharge tube between the electrodes had no noticeable influence on the onset and offset potentials. Under some conditions where the oscillator would not start the discharge it could be started with the aid of a small Tesla coil.

The offset potentials were closely reproducible, and those recorded were the lowest values obtained after several starts and stops, in accordance with the practice of other investigators.

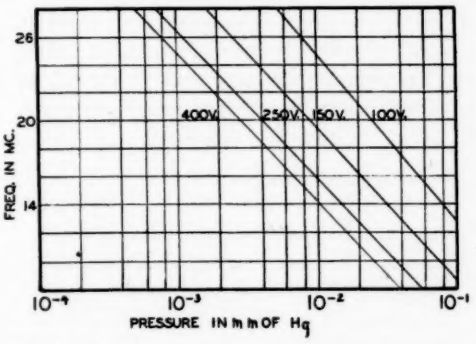


FIG. 6. Frequency *versus* pressure for several constant offset potentials, taken from Fig. 5.

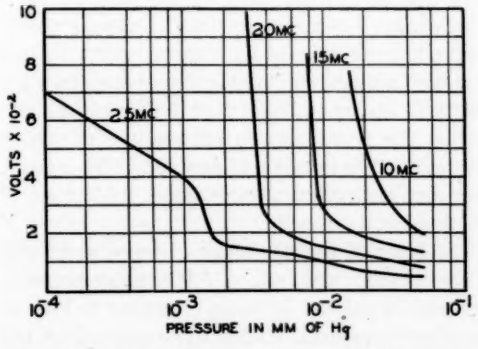


FIG. 5. Offset potentials *versus* pressures in dry air, using the three-electrode ring system with a spacing of 2.2 cm between the center ring and either of the external rings.

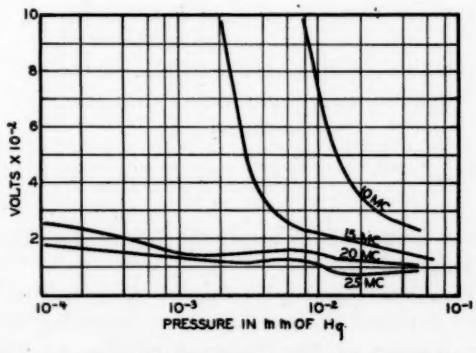


FIG. 7. Offset potential *versus* pressure in dry air, using the three-ring electrode system with a spacing of 4.75 cm.

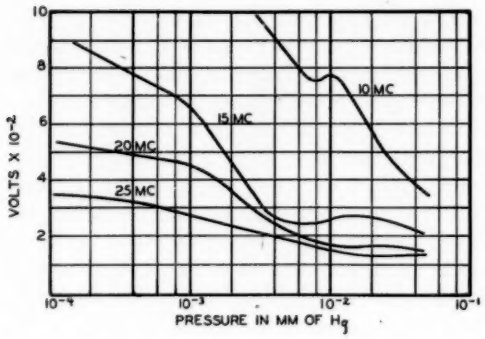


FIG. 8. Offset potential *versus* pressure in dry air, using the three-electrode ring system with a spacing of 15 cm.

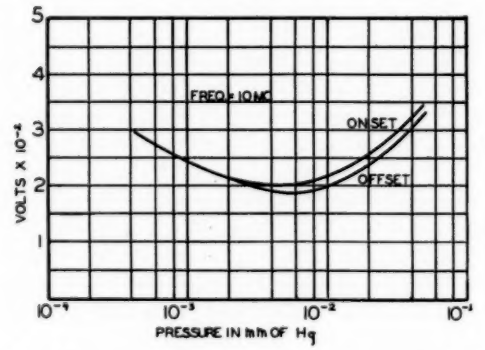


FIG. 9. Onset and offset potentials *versus* pressures in dry air at 10 megacycles/sec. using coil excitation.

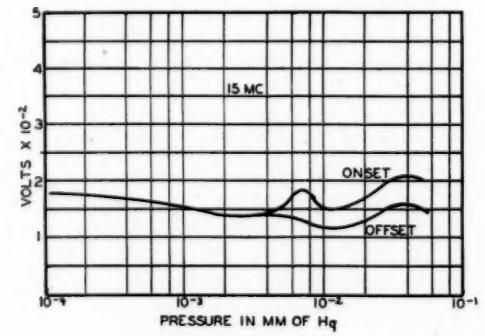


FIG. 10. Onset and offset potentials *versus* pressures in dry air at 15 megacycles/sec. using coil excitation.

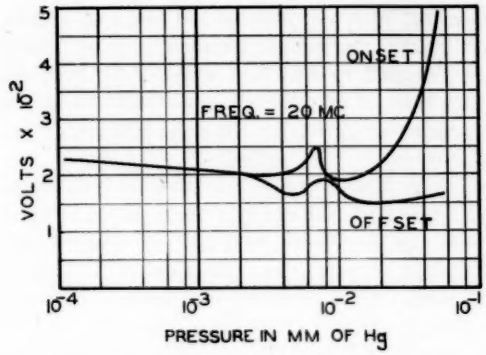


FIG. 11. Onset and offset potentials *versus* pressures in dry air at 20 megacycles/sec. using coil excitation.

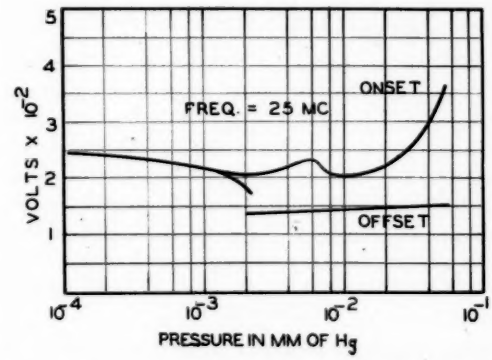


FIG. 12. Onset and offset potentials *versus* pressures in dry air at 25 megacycles/sec. using coil excitation.

These lowest values vary only slightly from the average values.

Double or even triple values of offset potential were obtained at some pressures, depending on the position of the discharge relative to the electrodes just before offset. Each of these values was distinct and reproducible. These three distinct types of discharge are shown in Fig. 3. When coil excitation was used several different types of discharge were again obtained. The three most frequently occurring are shown in Fig. 4.

In some pressure ranges the discharge would suddenly change from one type to another as the potential was varied. At the lower pressures the discharge became so faint that it was necessary to darken the room and to enclose the discharge tube in a shield in order to observe it. There was, however, always an abrupt onset and offset of visible glow detectable.

The color of the discharge changed markedly as the pressure was reduced. At high pressures it was pink or salmon in color, changing to blue at the lower pressures. Under some conditions the color varied in different parts of the discharge. No record was kept of these details except to note

that there was a rather sudden change in the appearance and color at about 1 to 2.5×10^{-3} mm. Below the pressure at which this change occurred only one type of discharge generally obtained, this being the centered type shown in Figs. 3 and 4.

Graphs of the data are presented in Figs. 5 to 12 and are largely self-explanatory.

In the regions where the offset potential was multi-valued the curves were drawn through the mean of the values.

5. DISCUSSION OF THE RESULTS

It seems most instructive to begin with the inspection of Fig. 5, where offset potentials for a visible glow discharge have been plotted against pressure at different frequencies. (As pointed out by earlier investigators,[2-4] onset potentials tend to be erratic; therefore they will not be discussed.) All four curves of Fig. 5 show a very steep rise of offset potential with decreasing pressure. Another general feature is that for a fixed peak potential the discharge ceases at lower pressures the higher the frequency. This pressure *versus* frequency relation is separately presented in Fig. 6 and suggests a logarithmic dependence $f=C\log p+D$. The slope of the straight lines of Fig. 6 is nearly the same, the constants have been determined from the graph for the 100-volt line such that

$$f=-11.7\log p+1.4.$$

It is of course realized that this equation applies only within the range covered by this investigation.

It is noteworthy that the 25-megacycle curve of Fig. 5 does not show the same marked rise as the other three do. In spite of careful outgassing of the tube walls and other precautions it is suspected that the pressures in the range of 10^{-3} mm to 10^{-4} mm were actually higher in the discharge region than indicated by the Pirani gauge. The fact that the discharge sometimes could not be restarted at low pressures after it had been stopped for more than a few seconds is further evidence that the pressure was being kept up by the discharge.

Figure 7 represents a curve similar to the one discussed above, but for larger electrode spacing (4.75 cm). The same general trends were observed, although for any given pressure and frequency the offset potentials were notably lower than was the case for closer electrode spacing. It seems out of place here to speculate on the different mechanisms which might account for the detailed structure of these curves. Nevertheless it is tempting to correlate the lower offset potentials at large electrode spacings with the findings of J. Thomson.[4] He pointed out that at large spacings the electrons can gain their greatest energy and thus ionize more efficiently than at the smaller spacings. In this latter case they quickly pass through the high field region into the space outside the grounded ring, where they are ineffective as ionizing agents.

The data obtained for 15-cm electrode spacing are presented in Fig. 8 and for coil excitation in Figs. 9 to 12. It should be noted that in the case of coil excitation the offset and onset potentials are far less dependent on frequency than in the case of ring electrodes. In addition a comparison between coil and ring excitation seems to indicate that for a given potential the discharge can be maintained at much lower pressures when coil excitation is used. No attempt is made to explain these phenomena.

One characteristic common to all curves showing irregularities is that these irregularities occurred at about the same pressures, namely from about 10^{-2} to 2.5×10^{-3} mm. This is undoubtedly connected with the fact that in this region the discharge has a tendency to jump from one type to another, thereby giving rise to multivalued offset potentials.

It should also be noted that at still lower pressures the curves become more regular with only one type of discharge prevailing. This low pressure discharge is quite different in appearance from that at higher pressure and is equivalent to what is referred to in the literature as a pure plasma discharge maintained by electrons undergoing Barkhausen-type oscillations about the center electrode.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 36, NUMBER 10 OCTOBER, 1946

Quantitative Tests for Off-Axis Parabolic Mirrors

L. G. SCHULZ*
Physics Department, University of Michigan, Ann Arbor, Michigan
(Received June 12, 1946)

Quantitative methods for examining and describing the surfaces of off-axis parabolic mirrors are given, together with practical suggestions for the most effective use of such mirrors. Included are tests with, and without, an optical flat. In each case the theory and experimental procedure are illustrated by application to a particular mirror. The quality of a mirror is given in terms of the minimum slit width that can be profitably used with it.

INTRODUCTION

THE well-known Foucault knife-edge test can be used only for the examination of symmetrical surfaces. Other methods, therefore, must be used for testing the unsymmetrical surface of off-axis parabolic mirrors. Several such methods will be described and evaluated; for the most part they are modifications and extensions of procedures previously given.[1-3] Off-axis parabolic mirrors as used in most optical arrangements have small "*f*-number," of the order *f*/3, consequently very well-defined shadow patterns will be formed when a quantitative knife-edge test is used.[4] These sharp shadows also make possible certain measurements made only with difficulty on the usual *f*/6 parabolic mirrors commonly used in telescopes.

The methods to be described were worked out while testing mirrors made by conventional polishing methods as well as those formed by deposition of evaporated metals on spherical surfaces. These mirrors were designed for use in infra-red spectrographs where attempts are being made to increase resolving power through improvement of optical parts. The obvious advantage of off-axis mirrors for many purposes together with the ease of manufacture by evaporation techniques should result in their increased use. It is felt that methods should be available which can be conveniently applied to test such mirrors.

EXPERIMENTAL ARRANGEMENT

Most of the tests require a slit-source of light and controlled knife-edge motion. The light source was conveniently made by cutting a slit in aluminum foil and then wrapping the foil around a six-volt light bulb about 6 mm in diameter. Several layers of paper were placed between the bulb and foil to diffuse the light and to reduce its intensity. In addition a Variac in the line was an appreciable aid in finding that intensity which gave optimum contrast in shadow patterns. The slit width was of the order 0.05 mm and was 2 mm in length. Heavy wire leads were soldered to the base of the bulb to act as a flexible support making it possible to place the source wherever desired for the different tests.

The knife-edge was formed by mounting a razor blade on a double screw motion, one motion parallel to the mirror axis and the other perpendicular. Each motion had a 0.5-mm pitch screw and could be read to 0.005 mm and estimated to 0.001 mm. To make full use of this sensitivity it was necessary to mount all parts, light source, knife-edge, mirror, and flat on a very solid base which was provided by a metal table built to support a Rayleigh interferometer. The legs of the table were separated from the floor by double boxes of sawdust. The procedure for using this equipment will be described with each of the methods which follow.

TESTING WITH A FLAT

The use of a flat makes possible a quick evaluation of a mirror in an arrangement ap-

* Now at Ohio State University, Columbus, Ohio.
[1] *Amateur Telescope Making* and *Amateur Telescope Making Advanced*, Sci. Am. (1937).
[2] John Strong, *Procedures in Experimental Physics* (Prentice-Hall, Inc., New York, 1945).
[3] E. Gaviola, J. Opt. Soc. Am. **26**, 163 (1936).
[4] Two definitions: *a*. The principal radius: The distance from the mirror axis to the remote edge of the mirror. This radius passes through the geometric center of the mirror. *b*. *f*-number of an off-axis mirror: The *f*-number of the smallest symmetrical mirror from which the off-axis mirror can be cut. One might call the usual ratio, aperture diameter/focal length, the effective *f*-number of the mirror.

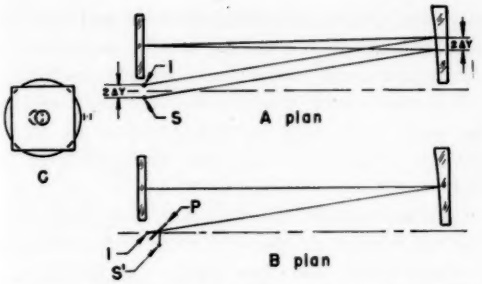


FIG. 1. Ray diagrams to show the origin of parallax. *A* and *B* are views from above and *C* is an end view of *A*. In *A* the source, *S*, is displaced from the axis by the amount ΔY; its image, *I*, is formed ΔY on the opposite side of the axis. The points of reflection on the mirror surface are separated by $2\Delta Y$, the same as the separation of *S* and *I*. In *B* the reflected ray returns on the path of the incident ray and there is no parallax. The use of a pellicle, *P*, makes it possible to place the source at *S'*.

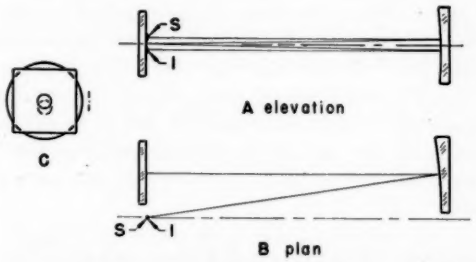


FIG. 2. An arrangement for avoiding the effects of parallax when testing for defects which are symmetrical to the mirror axis. *A*, as seen from the side; *B*, as seen from above; and *C*, as seen from the position of the tester. This is the most sensitive arrangement for detecting mirror defects.

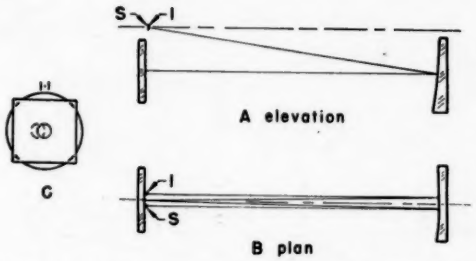


FIG. 3. An arrangement for minimizing the effect of defects symmetrical to the axis when the mirror is being used in an optical system employing a slit source. *A*, as seen from the side; *B*, as seen from above; and *C*, as seen from the end.

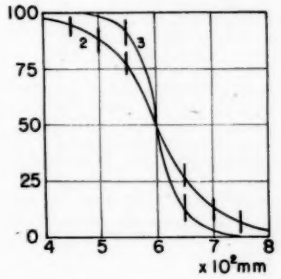


FIG. 4. A graph showing the rate at which a mirror surface darkened as a knife-edge was moved through the image. The vertical scale is the percentage of the surface illuminated, the horizontal scale the motion of the knife-edge, left to right. Curve 2 is for the arrangement of Fig. 2; Curve 3 for the arrangement of Fig. 3.

proximating that in which it is likely to be used. Unlike the case when testing symmetric parabolic mirrors, it is not necessary to have a hole through the flat, since it, like the mirror is displaced to the side of the axis. Unfortunately, unless the source and its image are superimposed, the two reflections which occur on the mirror being tested are not on the same surface element but on two different elements whose separation is directly proportional to the mirror and flat separation as is shown in Fig. 1A. A knife-edge test on the image, *I*, will not reveal the nature of any one surface element but rather an average of two separated elements. A pellicle can be used to effectively superimpose the source and its image as in Fig. 1B, but such a complication can be avoided by proper geometrical arrangement; Fig. 2 shows the relative position of mirror, flat, source, and knife-edge. In particular the source and knife-edge are on a common vertical line perpendicular to the axis; any parallax will, therefore, be in a vertical direction. Because defects are in general symmetrical relative to the axis even for off-axis mirrors, it is possible to examine the smallest arc defects on the mirror. As a result of the double reflection, defects will be magnified. Obviously this is a poor arrangement for using the mirror, but no worse than having the slit and its image side by side as is usually done, Fig. 1A.

The best arrangement for *using* an off-axis mirror with a slit source is that shown in Fig. 3, where the slit and its image are side by side and parallel to the principal radius. Defects symmetrical to the axis will produce decidedly less image broadening than was the case with the first arrangement. Naturally, this is an insensitive arrangement for *testing* the mirror. In an infrared spectrograph it is convenient to have the

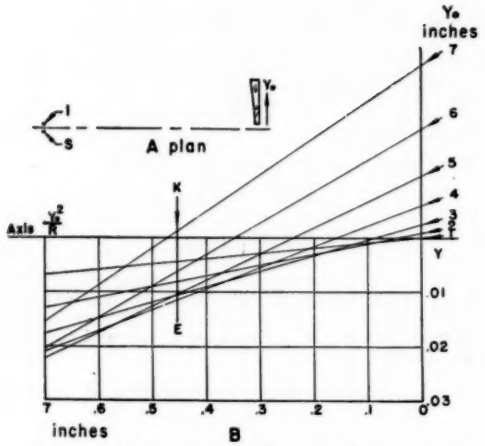


FIG. 5. In A is shown the relative positions of mirror, light source, and image for testing without a flat. The source and its image are approximately at the center of curvature of the mirror. In B is shown the manner in which rays pass through the image. Y_0 is the point on the mirror from which the ray has been reflected; X_e of Eq. (6) is measured along the Y_0^2/R scale. The line KE indicates the path followed by the knife-edge in obtaining the shadow pattern of Fig. 8. All dimensions are in inches.

prism axis vertical; consequently, the slits must also be vertical.

The experimental procedure was to adjust the knife-edge motion parallel to the axis until the mirror darkened most evenly on moving the knife-edge in perpendicular to the axis. Then the extent of darkening was recorded graphically as a function of this perpendicular motion. Figure 4 shows the result for a mirror formed by evaporation: Curve 2 is taken with the arrangement of Fig. 2; Curve 3 with that of Fig. 3. A slit width of 0.02 mm might be used with profit in the first case and one of 0.01 mm with that for arrangement of Fig. 3. Later, in Fig. 11, it will be shown geometrically why there is less broadening in one case than the other.

A direct method of measuring the extent of the parallax and at the same time an aid in aligning the optical parts of the system was to place a small circular loop of wire directly in front of and at the center of the mirror surface. It, together with its shadow, formed two dark circles separated by the extent of the parallax. In each of Figs. 1–3C the actual wire loop is shown solid whereas its shadow is broken. The position of shadows for this test, as well as the other tests, was read on a cardboard scale placed along the principal radius of the mirror. Inch marks were located by deep notches, half-inch ones by smaller notches. On the opposite edge were placed quarter-inch cuts. Diffraction effects made these cuts very easy to see and in general it was possible to locate shadow edges to better than one-eighth of an inch.

The Ronchi test was found to give very obvious qualitative evidence of minor local defects. For this test a 200-mesh screen was fastened to the knife-edge, one set of wires parallel to the edge. The pattern appearing on the mirror when it was viewed through the screen depended on the relative position of the optical parts. For example, using the arrangement of Fig. 2, as the knife-edge is moved back from the focus position, the vertical lines remain continuously visible, whereas the horizontal ones because of parallax at times did not appear. When the horizontal lines were sharpest for the mirror as a whole, defective areas would be outlined by a fuzzy appearance of the horizontal lines. With the source to one side of the knife-edge as in Fig. 3 the effect of parallax appears on the vertical lines.

THEORY OF A QUANTITATIVE METHOD NOT REQUIRING AN OPTICAL FLAT

The method and ideas of Gaviola for the quantitative use of the knife-edge test were extended.[3] It was found that in general a defect in the mirror figure appears most conspicuous when seen from the average center of curvature of the region surrounding the defect. Depending on the direction of the motion of the knife-edge, this is equivalent to putting the defect in the region of last darkening or first darkening, the contrast being higher for last darkening. In an effort to gain an over-all test of a mirror, experience showed it was easy and convenient to divide its surface into regions and examine each at its average center of curvature. Considerations of the properties of a parabola suggested a method. Thus, if the equation of the parabola is given in the form

$$X = Y^2/2R, \tag{1}$$

the slope of the normal at the point X_0Y_0 is

$$dY/dX = -Y_0/R. \tag{2}$$

The equation of the normal through X_0Y_0 is

$$Y-Y_0=-(Y_0/R)[X-(Y_0^2/2R)], \quad (3)$$

and its intersection on the axis is

$$X=R+(Y_0^2/2R). \quad (4)$$

If a fixed light source is placed on the axis a distance R from an on-axis symmetrical parabolic mirror, the image of this source will be spread over a range

$$\Delta X=Y'^2/R, \quad (5)$$

where Y' is the mirror radius.

Figure 5 shows how the reflected rays cross the mirror axis. A ray from any point Y_0 on the principal radius of the mirror crosses the axis at Y_0^2/R with a slope of Y_0/R. From the geometry of the figure the envelope of this family of straight lines is given by the parametric equations

$$X_e=3Y_0^2/R \quad (6)$$

$$Y_e=(Y_0/R)[X_e-(Y_0^2/R)]. \quad (7)$$

Putting (6) into (7),

$$Y_e=2Y_0^3/R^2.$$

This envelope represents the locus of positions giving the focus of points Y_0 on the principal radius of the mirror. Also, if the knife-edge moves in a downward direction the envelope is the locus of knife-edge positions for last darkening. Some definite value for X_e may be put into Eq. (7) which is rewritten as

$$Y=(Y_0/R)[X_e-(Y_0^2/R)]. \quad (8)$$

Y is now the displacement of the knife-edge perpendicular to the axis which will just cut off the rays from Y_0, in general double valued.

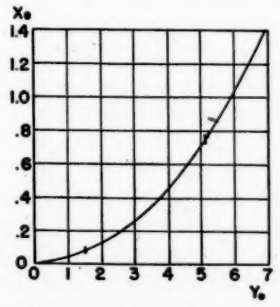


FIG. 6. Graph showing the position of the knife-edge to obtain last darkening at point Y_0 on the mirror surface. The two experimental points at Y_0 equal to 1.5 inches and 5.2 inches were used to locate the zero point on the parallel motion of the knife-edge. Both scales are in inches.

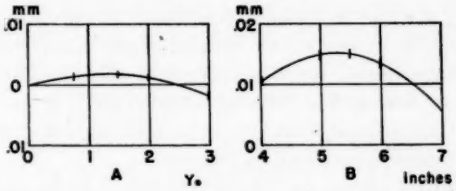


FIG. 7. Two graphs showing the manner in which the position of last darkening is located from the experimental data. The two values are indicated in Fig. 6.

In words, when the plane of the knife-edge is $R+X_e$ from the mirror as measured along the mirror axis, then the shadow edge will be at Y_0 from the axis when the perpendicular motion of the knife-edge has carried it Y distance over the axis.

APPLICATION OF THE THEORY

The procedure in using the above method is best illustrated by considering a mirror with a rather obvious defect. The mirror chosen had been parabolized by evaporation of aluminum onto a spherical mirror. It was 6 inches in diameter, its radius of curvature, R, equal to 104 inches, and its axis one-half inch outside the periphery. Using this value of R, a graph was made of Eq. (6) and shown in Fig. 6. Since X_e is least sensitive to changes in Y_e for last darkening near the inside edge, the parallel motion of the knife-edge was set to give last darkening in this region and a graph, Fig. 7A, made of the perpendicular motion of the knife-edge as a function of the shadow position on the mirror. The value Y_0 equals 1.5 inches from Fig. 7A is placed on Fig. 6. By this process one has established the X_e zero on the instrumental scale. This can be further checked by finding last darkening near the outer edge, Fig. 7B, which on the instrumental scale gave X_e equal to 0.75 inch for a Y_0 of 5.2 inches, a value which fits well on Fig. 6. To get the central defect at Y_0 equal to 3.5 inches into sharp relief, X_e was set equal to 4.52 inches and a complete graph made of Eq. (8) and shown in Fig. 8A. It is evident that the figure of the mirror as a whole is good except for a hump at 3.5 inches and a turned down outer edge. It should be understood that this center defect can be found with any setting of X_e as it appears in

Fig. 9A and B. It is at the center of curvature, however, that a defect is most easily discovered and measured, for it is there that the simplified set-up is most like the test in which a flat is used, in that defects show up on a relatively evenly illuminated surrounding region.

Mirror testing by this method might be divided into two parts: testing the figure as a whole and detailed examination of defects. If the mirror has many defects, it is advisable to use a rather wide slit source to wash out these defects when testing for the figure as a whole. On the other hand, a very narrow source must be used for examining local areas, with special care given to insure that the source and knife-edge are parallel.

EVALUATION OF A MIRROR

It has been the practice to measure the quality of a mirror by the trueness of its figure, but this is not necessarily a test of its performance. Rather it is the relative slope of each surface element which is significant. Consider for example the mirror described by curves in Figs. 8 and 9. The contour of this mirror along the principal radius is that given in Fig. 10 where the horizontal base line represents the original spherical surface of the mirror before evaporation, and the vertical scale the thickness of the metal deposited on it. The continuous curve, then, is the metal thickness as a function of the distance from the mirror axis, and is given by the equation

$$T = Y_0^2(Y'^2 - Y_0^2)/8R^3.$$

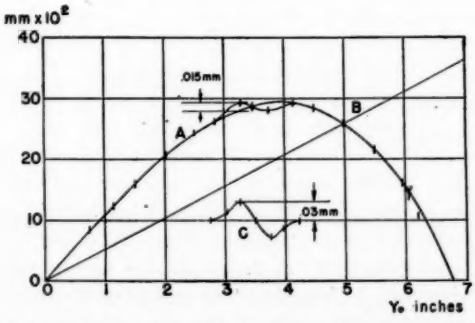


FIG. 8. The curve A is the shadow pattern which should result when the knife-edge is moved through the image as shown in Fig. 5. The double curve is drawn to fit the experimental points when they deviate from the theoretical curve. Curve B is the shadow pattern for the spherical mirror before it was parabolized. In C is shown the shadow pattern obtained when the defect between Y_0 equals 3 and 4 inches is examined with a flat, arrangement Fig. 2.

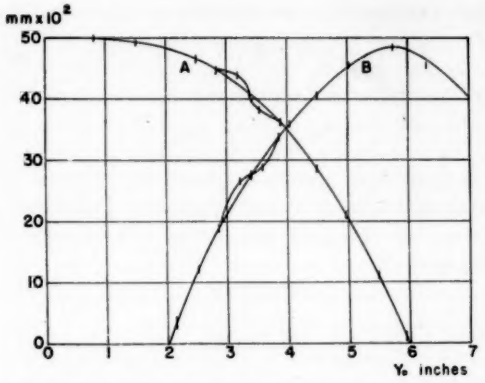


FIG. 9. Shadow patterns when the knife-edge is placed for last darkening at Y_0 equal to zero and to 5.7 inches. The smooth curves are the theoretical patterns, and the vertical lines the experimental points.

Here Y_0 is measured along the principal radius and Y' is its maximum value. R is the radius of curvature of the mirror. The 6-inch mirror with the axis $\frac{1}{2}$ inch outside its periphery extends from Y_0 equal to $\frac{1}{2}$ inch to Y_0 equal to $6\frac{1}{2}$ inches. With R equal to 104 inches the maximum thickness at Y_0 equal to 5 inches is 17,000A. The broken lines between Y_0 equal to 3 inches and 4 inches show the defect of Figs. 8 and 9. At best, Fig. 10 indicates the extent and magnitude of this defect but not at all its effect on the mirror's performance. This can, however, be obtained directly from Fig. 8: From curve A the arc of radius $3\frac{1}{4}$ inches and width about $\frac{1}{4}$ inch will deflect light (0.015/2) mm out of the true image when the mirror is used as the reflector of a telescope. The arc at $3\frac{3}{4}$ inches will broaden the image in the opposite direction for a total broadening of 0.015 mm. From curve C when the mirror is used with a flat, arrangement Fig. 2, the same defect will deflect light 0.03 mm outside the image in both directions, or a total slit widening of 0.06 mm. If now the optical system is changed to that of Fig. 3, then the part of the defect near the principal radius will produce no image widening, and at worst the part of the defect near the edges a widening of about 0.05 mm or an average of 0.025 mm. Thus the deleterious effect is reduced to less than one-half. The schematic drawings B and C of Fig. 11 show geometrically that with the axis farther and farther from the mirror the effect of defects becomes progressively less.

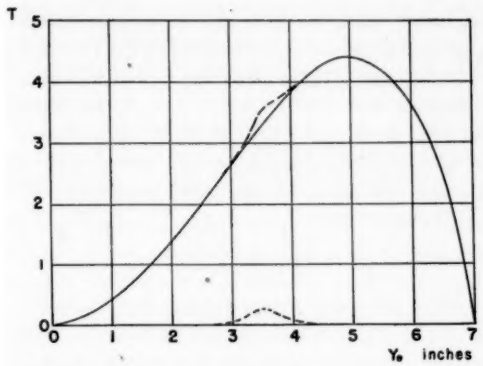


FIG. 10. The relative thickness of metal, T, on a parabolized mirror as a function of the distance Y_0 along the principal radius. The horizontal base line represents the original spherical surface, the solid curve the theoretically correct deposit, and the broken curve the actual thickness over the defective area of Figs. 8 and 9.

For a symmetrical mirror the arc with a radius at the periphery contributes most in determining the quality of the image, whereas for an off-axis mirror the arc with a radius slightly beyond the center of the mirror is most effective. The process of weighing and averaging the contributions of all the surface elements is most quickly done by the method already illustrated in Fig. 4 where the quantitative result in terms of slit width was obtained directly. This feature of testing with a flat might seem to be its greatest advantage, but unfortunately the eye needs to be trained to become a trustworthy integrating instrument if the defects are sharp and deep. Perhaps a photo-tube detector would be better.

The mirror used for illustration through most of this discussion was not a good one; it was selected because its obvious defect could be easily shown on a small graph. It should be evident that defects of about 1/20 its magnitude can be detected and those as small as $\frac{1}{10}$ evaluated quantitatively, or in terms of figure variation 0.02 and 0.01 wave-lengths, respectively. The evaporation technique used to parabolize a spherical mirror followed the general procedure given by Strong[2] but with greatly improved geometry. In particular the filament was a high pitched helix of 30 loops bent into a circle of a diameter 10 inches. It was possible to use the diaphram as calculated, no empirical alterations being necessary. Evaporations were carried out in an 18-inch bell jar. A small electric motor used to turn the mirror was contained within the vacuum.

COMPARISON OF THE METHODS

The advantages of using a flat for testing are clearly shown by the curves of Fig. 8. Curve 8A is taken with the arrangement of Fig. 5A, not using a flat; curve 8C using a flat as in Fig. 2. The same defect at Y_0 equal to 3.5 inches appears twice as large in C as in A, and also is more easily measured in C because of the sharper contrast. Simple geometric considerations show that if the flat is placed approximately at the focal distance from the mirror, the linear deviation of light from the defective area is twice that obtained from the arrangement without a flat. In addition, the contrast is sharper because the angular deviation is also twice as great. By the simple expedient of increasing the mirror and flat separation, the linear deviation may be arbitrarily increased. This, however, is a doubtful gain because a diagonal will be needed to bring the image to the side where it can be examined without the head of the observer intercepting light to and from the flat. It was found more advantageous in practice to bring the flat as close to the mirror as possible to avoid air turbulance created by the face and hands of the experimenter.

In making comparisons it must not be overlooked that the mounting of the flat must be as rigid as the rest of the optical system. Also, there

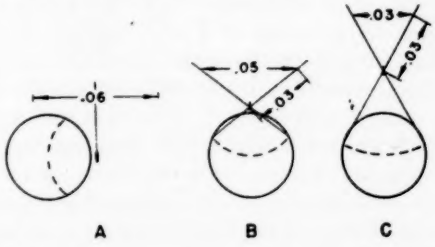


FIG. 11. In each figure the circle represents the 6-inch mirror with an arc defect through the geometric center, as in Figs. 8 and 9. In A, arrangement of Fig. 2, the slit is perpendicular to the principal radius and on the axis which is $\frac{1}{2}$ inch outside the mirror edge. In B, arrangement of Fig. 3, the ends of the defect would at worse broaden the image by 0.05 mm or a mean of about 0.02 mm. In C the axis is outside the mirror a distance equal to the radius. Assuming a defect of the same magnitude as that in A and B the broadening in C will be reduced to a mean of about 0.015 mm.

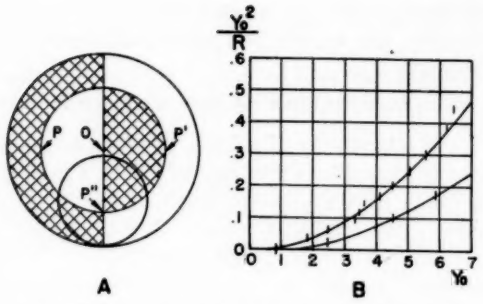


FIG. 12. *A* shows a typical "doughnut" shadow pattern on a symmetrical on-axis parabolic mirror. The smaller circle represents a small off-axis mirror cut from the larger one. B is a graph showing completed parabolization for the 6-inch mirror of Fig. 8. The lower curve is for half-completion. Both scales of B are in inches.

is some time required to adjust each element of the optical system. Obviously, the defects of the flat must be of an order lower than those of the mirror being tested. It is suggested that this set-up using a perfect off-axis parabolic mirror be used to test membrane flats which cannot be checked against a standard by contact methods.

A QUICK TEST FOR THE EXTENT OF PARABOLIZATION

When mirrors are made by the evaporation technique, it is convenient to have a quick method of determining the extent of parabolization and at the same time obtain a quick survey of the entire surface. Such a method is suggested by consideration of Eqs. (1) to (4) which show that an arc a distance Y from the axis has its center of curvature at $R+(Y^2/2R)$. When the image formed by the entire surface of a symmetrical on-axis mirror is examined anywhere in the range R to $R+(Y^2/R)$ (fixed light source) the shadow pattern is that usually designated as the "doughnut" pattern. The Foucault test is a method of locating points PP' of Fig. 12A by means of diaphrams. The equivalent process for an off-axis mirror is to locate the point of reversal, P''. If the mirror is of small f-number this can be done very accurately because of the sharp shadows. This method was especially useful in examining the outer edges of mirrors. Figure 12B shows the results for the 6-inch mirror in Fig. 8. Although this method is quick, it is not suitable for the final test because it is insensitive for measuring local defects and the results are not directly interpretable in either figure defect or performance defect.

CONCLUSIONS

Although the professional optician has his own methods to guide him in shaping a mirror surface, these methods are in general unfamiliar to, or unusable by, the average physicist. The purpose here has been to describe such methods as can be conveniently performed with equipment generally available, thus enabling the instrument designer to learn for himself the quality of parts composing an optical system. Special attention was given to slit sources since point sources are rarely used other than in astronomy. Because of the peculiar properties and advantages of off-axis parabolic mirrors it was felt that the information here given would be helpful to others. Although the purpose has been to describe arrangements for testing purposes, Fig. 2, attention was also called to an arrangement, Fig. 3, for using the mirror to greatest advantage. The effects of arc defects of an off-axis mirror will be minimized by placing the mirror with its principal axis parallel to the slit system; the farther the axis from the center of the mirror the greater will be the advantage of this arrangement, Fig. 11B and C. It should be mentioned that the test method without a flat can be used equally well on an on-axis mirror.

The mirrors tested were loaned by Professors Randall and Williams, other optical parts by Professor Meyer and Paul Weyrich. The author wishes to thank these people for the use of the equipment as well as suggestions. The work was supported in part by funds from the Evaporated Metal Films Corporation, administered through the Department of Engineering Research of the University of Michigan.

Color Vision and Color Blindness: A Mechanism in Terms of Modern Evidence*

THEODORE LOUIS JAHN
Zoological Laboratories, State University of Iowa, Iowa City, Iowa
(Received June 12, 1946)

GRANIT[1-3] has demonstrated by means of microelectrodes placed on nerve fibers in the vertebrate retina that there are present several types of receptor units. The most numerous type has a wide spectral sensitivity curve almost identical with the human luminosity curve and is called a "dominator"; other types have narrow spectral sensitivity curves and are called "modulators." These experiments form a basis for his dominator-modulator theory of color vision.

In recent publications[2-6] there are discussions of the cause of protanopia (red blindness) and deuteranopia (green blindness) in terms of Granit's theory. Granit[2] assumes that in deuteranopia both "red" and "green" modulators are missing and that the dominator alone is responsible for the luminosity curve (as is also assumed for the normal eye). This explanation is inadequate in that it considers the deuteranope to be mono- rather than dichromatic, and this is contrary to all available data.[6-9] The commonly accepted explanation of deuteranopia which assumes an absence of the green sensation is not subject to this criticism. The commonly accepted explanation is not at variance with Granit's data but only with the theory as expressed above. Willmer[5] proposed an alternative theory which has been the subject of considerable criticism.[10]

Granit's data may be combined with that concerning the synaptic connections in the retina[11] to provide the basis for a very simple explanation of the mechanism of color vision and its deficiencies. Some fibers in the optic nerve are connected through multisynaptic ganglion cells and bipolar neurones to rods and cones; others are connected through midget ganglion and bipolar cells to single cones. Some cones are connected to both types of nerve fibers. These pathways are shown diagrammatically in Fig. 1.

These connections form a basis for Granit's dominators and modulators provided we consider these terms to refer specifically to ganglionic axons and not to types of cones. On this basis fibers of the multisynaptic ganglion cells become the dominators.[3] The modulators are of six or seven types which for convenience are considered as three groups, namely, red, green, and blue. The algebraic summation of the modulator sensitivity curves yields that of the

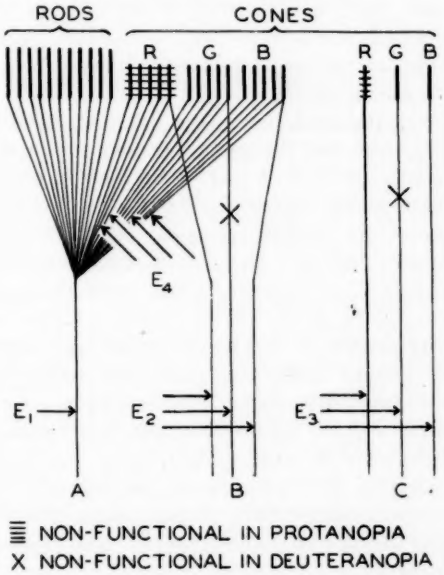


FIG. 1.

* Aided by grant from the Rockefeller Foundation for research on the Physiology of the Normal Cell.

[1] R. Granit, J. Opt. Soc. Am. **31**, 570–580 (1941).
[2] R. Granit, Nature **151**, 11–14 (1943).
[3] R. Granit, Nature **155**, 711–713(1945); J. Neurophysiol. **8**, 195–210 (1945).
[4] H. Hartridge, J. Physiol. **103**, 10*P* (1945).
[5] E. N. Willmer, Nature **151**, 213–215 (1943).
[6] D. B. Judd, J. Research Nat. Bur. Stand. **33**, 407–437 (1944).
[7] G. A. Fry, J. Opt. Soc. Am. **34**, 159–166 (1944).
[8] D. B. Judd, J. Opt. Soc. Am. **33**, 294–307 (1943).
[9] F. H. G. Pitt, Med. Res. Council Gr. Brit., Spec. Rep. No. 200 (1935); Proc. Roy. Soc. **B132**, 101–117 (1944).
[10] F. W. Edridge-Green, Nature **151**, 422 (1943); H. Hartridge, *ibid.* **151**, 422 (1943); E. N. Willmer, *ibid.* **151**, 632 (1943); K. J. W. Craik, *ibid.* **151**, 727–728 (1943); A. H. S. Holbourn, *ibid.* **152**, 190 (1943); H. Hartridge, *ibid.* **152**, 190 (1943); E. N. Willmer, *ibid.* **152**, 191 (1943).
[11] S. L. Polyak, *The Retina* (Chicago University Press, Chicago, 1941); Biol. Symposia **7**, 193–202 (1942).

dominator.[8] Therefore, we may consider the dominator fiber to be connected to the receptors as in *A* (Fig. 1) and the recording electrode in position E_1. The modulator fibers may be attached to individual cones as at *C* or to some of the cones which are attached to the dominator as at *B*, and modulator responses are recorded at E_2 and E_3. A sensitivity curve identical with that of modulator could also be recorded from any single preganglionic fiber connected as at E_4. We may assume that the luminosity curve is determined by the total effect of all dominators and modulators (rather than of the dominator alone, as proposed by Granit) and that hue discrimination is determined by the modulators alone.

On the basis of the above theory protanopia and deuteranopia can easily be explained by assuming specific deficiencies. Protanopia is apparently a deficiency of the "red" cones attached to both multi- and monosynaptic neurones, i.e., to both dominator and modulator fibers. The absence of functional "red" modulators would both limit red-green hue discrimination and slightly lower the luminosity curve at the red end of the spectrum, and the absence of functional "red" cones among those connected to the dominator would result in the considerable shortening of the luminosity curve which is known to occur.[6–9]

In deuteranopia red-green hue discrimination is very low, but the spectral sensitivity curve is almost normal. This can be explained as a deficiency in the "green" modulator units (midget bipolars, or midget ganglion cells, or their attached "green" cones) so that the pathways marked *X* in Fig. 1 become non-functional, either in the retina or between the retina and higher centers of the brain. Pathways between the "green" cones and their dominator fiber must remain functional. This assumption would explain a great decrease in hue discrimination accompanied by only a slight change in spectral sensitivity, such as is known to occur.[6–9] The fact that nerve fiber degeneration produces deuteranopia but not protanopia[12] also supports this theory.

In tritanopia (blue-blindness) the available evidence[9] indicates that the visual spectrum is almost normal. Therefore, the inability to distinguish color is probably caused by a deficiency in the "blue" modulator pathways (cf., however, reference 15).

Pickford[13] described a case of one-eye monochromatic (green sensitive) color blindness (following an accident) in which the visibility curve was apparently normal, and he pointed out that this could not be explained by any of the current two- or three-color theories. This phenomenon can be explained on the present theory by assuming a deficiency in the "red" and "blue" modulator pathways similar to that assumed in the "green" pathway for deuteranopia or in the "blue" pathway for tritanopia.

The fact that very small objects, especially at the fovea, may be seen by the normal eye as with partial tritanopic rather than with trichromatic vision has been the subject of much discussion.[14,15] If the "blue" modulators are assumed to be much less numerous than the "red" and "green" ones such a result is to be expected on a statistical basis as the number of excited cones is reduced.

The difficulties of explaining all color vision data on the basis of the more modern versions of the König trichromatic "excitation" curves[16] were pointed out by Hecht[17] who proposed to substitute a new set of curves. These new curves differ from previous ones in that the three spectral sensitivities are only slightly different from each other. By assuming the existence of these curves, Hecht was able to derive equations which fit the data for color mixture, luminosity, saturation, estimation of colorimetric purity, hue discrimination, and deuteranopia. Mandelbaum and Mintz[18] demonstrated that dark adaptation curves could also be predicted on this assumption. The only basic data on color phenomena

[12] H. Kollner, *Die Störungen des Farbensinnes* (Karger, Berlin, 1912).

[13] R. W. Pickford, Nature **155**, 336–337 (1945).

[14] E. N. Willmer, Nature **153**, 774 (1944); W. S. Stiles, *ibid.* **155**, 177 (1945); L. C. Thomson, *ibid.* **155**, 177 (1945); M. H. Pirenne, *ibid.* **155**, 177–178 (1945); H. Hartridge, *ibid.* **155**, 391–392, 657–662 (1945); E. N. Willmer and W. D. Wright, *ibid.* **156**, 119–121 (1945).

[15] H. Piéron, Nature **157**, 106 (1946).

[16] L. T. Troland, J. Opt. Soc. Am. **6**, 527–596 (1922).

[17] S. Hecht, J. Opt. Soc. Am. **20**, 231–270 (1930); J. Opt. Soc. Am. **21**, 615–639 (1931); Murchison, *Handbook of General Experimental Psychology* (Clark University Press, Worcester, Massachusetts, 1934), pp. 781–821.

[18] J. Mandelbaum and E. U. Mintz, Am. J. Ophthal. **24**, 1241–1253 (1941).

which Hecht's theory does not explain are those pertaining to protanopia.

The demonstration by Granit of the existence of modulators with widely separated sensitivity bands is obviously contrary to the assumption that Hecht's curves represent the sensitivities of individual cones. However, Hecht's theory is the only one which gives an adequate description of most color data (all except protanopia). If we consider Hecht's curves in terms of sensation[19] rather than in terms of cones, all contradictions are removed and even protanopia can be explained. Hecht's red, green, and blue curves may be considered as the curves for the sensation produced by the combined action of the dominators plus any one of the three types of modulators. Hecht's curves then become the result of the fused action of both dominators and modulators as measured at some higher center in the central nervous system. The mechanism of this assumed fusion is not necessarily different from that of the fusion of the effect of any group of modulators to give the sensation of white or of some band of color. The explanation of protanopia offered under the Granit theory then automatically covers the phenomenon under the Hecht theory.

The above combination of modern theories of color vision is apparently capable of explaining all of the basic types of data concerning color vision and the various types of color blindness. Protanomaly and deuteranomaly are examples of partial protanopia and deuteranopia. The two types of total color blindness (achromatopsia), as pointed out by Granit,[3] may be caused by a complete absence of either cone or of modulator function.

[19] G. L. Walls, *The Vertebrate Eye* (Cranbrook Institute of Science, 1942).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 36, NUMBER 10 OCTOBER, 1946

Rapid Response Thermopiles*

LOUIS HARRIS
Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received June 17, 1946)

A technique has been developed for the evaporation of bismuth and of antimony upon thin cellulose nitrate so that the metals have comparatively low electrical resistivities and comparatively high thermal e.m.f.'s. A procedure has been developed for making bismuth-antimony radiation thermopiles in which units of 50 thermal junctions have a receiving area of ~0.11 cm² and a resistance of ~70 ohms. These units are called "folded thermopiles." The response of several folded thermopiles of different design operated at atmospheric pressure has been studied over a range of frequencies. The "folded" thermopiles are faster than any thermopiles reported heretofore, yet the response is greater for all frequencies above 5 cycles per second than for evaporated thermocouples operated in a vacuum. The thermopiles are rugged, relatively free of microphonics, and show very little noise in addition to "Johnson noise."

INTRODUCTION

THE speed of response of a radiation thermocouple is greater (1) the smaller the thermal mass and (2) the greater the heat loss. Sputtered[1] and evaporated[2,3] thermocouples with speeds high compared to wire thermocouples have been obtained by making the thermal mass as small as possible. Further increase in speed of response, through decreased thermal mass, awaits the discovery of new materials and/or methods of obtaining thinner deposits of metals having more favorable electrical and thermal properties. The gain in speed through increased heat loss would appear to have little advantage since the μ volt/μ watt/cm² would decrease accordingly. However, it has been possible to achieve high heat loss and therefore higher speed of response

* The work described here was carried out in whole or in part under contracts OEMsr-126 and OEMsr 1147 between the Office of Scientific Research and Development and the Massachusetts Institute of Technology.

[1] L. Harris and A. C. Scholp, J. Opt. Soc. Am. **30**, 519 (1940).

[2] A. Stockfleth, M.I.T. Ph.D. Thesis (November, 1942).

[3] L. C. Roess and E. N. Dacus, Rev. Sci. Inst. **16**, 164 (1945).

TABLE I.

Metal	Thicknesses (microns)	Resistivity (ohm-cm) average	Resistivity (ohm-cm) limits	Thermal e.m.f. vs. copper (microvolts per degree C) average	Thermal e.m.f. vs. copper (microvolts per degree C) limits
Bismuth	0.73 to 1.58	168×10^{-6}	$131\text{–}212\times10^{-6}$	−57.5	−55.1 to −59.0
	1.64 to 2.14	159×10^{-6}		−56.1	
	2.05 to 3.15	130×10^{-6}		−56.0	
Antimony	0.48 to 1.66	95×10^{-6}	$60\text{–}119\times10^{-6}$	35.5	33.1 to 38.6

and also comparatively high μ volt/μ watt/cm^2. This was accomplished by using thicker metal deposits, so that most of the heat loss occurred by metallic conduction, and by crowding many thermocouples into a small receiving area.

We shall describe in detail the construction and response tests of a fifty-junction thermopile having a receiving area 0.25 cm$\times$0.44 cm. Each element is a single junction thermocouple formed by evaporating bismuth and antimony on opposite sides of a thin *folded* cellulose nitrate film. The overlapping junction of the two metals is at the fold and is the "hot junction." Fifty elements (0.25 cm wide) are assembled in a miniature press so that the thermojunctions, one above the other, form the receiving area. Electrical contact among the elements is made mechanically at approximately 0.030 cm behind the fold. The contact lines form the cold junctions. Figure 3 shows the structure of the thermopiles whose method of manufacture will be described in the next section.

CONSTRUCTION OF THERMOPILES

Method of "Evaporating" the Metals

Bismuth and antimony were deposited by "evaporation" on to cellulose nitrate by the following procedure. A weighed quantity of the pure metal (previously cast in vacuum) was evaporated from a "boat"; the current through the boat was increased as the evaporation proceeded so that the boat was at red (for antimony) or yellow (for bismuth) heat at the end of the evaporating process. Alongside the boat containing the metal to be evaporated was another boat (empty) maintained at red (for antimony) or yellow (for bismuth) heat, during the evaporation process. The evaporation of each metal was carried out stepwise, so that about twenty-five minutes elapsed for a complete evaporation. A vacuum of approximately 10^{-4} mm Hg was maintained during the evaporation process. Tantalum boats were used for the bismuth evaporation; molybdenum boats were used for the antimony evaporation. The progress of the evaporation was followed with a "meter." The meter was situated in the bell jar directly above the boat from which the metal was evaporated.

The detailed procedure used for the evaporation of antimony has been described.[4]

We found, in agreement with Bussem, Gross, and Hermann,[5] that it was necessary to form the bismuth deposit at an elevated temperature, in order to obtain electrical resistivity not greatly in excess of the macroscopic value. We obtained the elevated temperature by radiation, by operating the empty boat adjacent to the bismuth sample boat at "yellow heat." Films of cellulose nitrate were occasionly decomposed by the high temperature maintained.

Measurements of Electrical Properties of Evaporated Bismuth and Antimony

Early experiments showed that the amount of metal deposited and the resistivity of the deposited metal varied with the evaporation technique as well as with the material upon which the metal was deposited. However, the results given below are for the metals deposited upon thin films of cellulose nitrate. The cellulose nitrate films were mounted on mica frames to facilitate handling. The films were placed in the bell jar, in horizontal position, about 15 cm above the boats. After completion of an evaporation the films were removed from the bell jar, leads were soldered to two ends of the evaporated metal with Wood's metal, and the resistance of each strip was determined. A strip about 2.2 cm$\times$1.1 cm of the

[4] L. Harris, J. App. Phys. **17**, 757 (1946).
[5] W. Bussem, F. Gross, and K. Hermann, Zeits. f. Physik **64**, 537 (1930).

metal with its backing was then cut out of the mica frame and weighed. The metal with its backing was then mounted on a piece of mica and the thermal e.m.f. *vs.* copper was measured. A strip of the cellulose nitrate used had been weighed previously so that the resistivity of the metals could be calculated. (The mass of the cellulose nitrate was about 0.35×10^{-3} gram while the masses of the metals varied from 0.7×10^{-3} gram to 4×10^{-3} gram. Massive density was assumed for the metals.)

The results obtained, using the technique described above, are found in Table I.

The bismuth deposits had a dull light-grey appearance; the antimony deposits were a brilliant steel-grey and showed a crystalline pattern to the unaided eye.

Burger and Van Cittert[6] obtained a thermal e.m.f. for evaporated bismuth-antimony of 75 μv/°C[9]. The values obtained here are about 25 percent greater.

The lowest resistivity measurements for evaporated bismuth are those reported by Bussem, Gross, and Hermann[5]; their value of 160×10^{-6} ohm-cm is nearly the same as ours. Each value is about 30 percent greater than that of the massive metal.

The resistivity of evaporated antimony obtained here is about $\frac{1}{10}$ of that reported by Stockfleth.[2] The resistivity for evaporated antimony is still about $2\frac{1}{2}$ times greater than that of the massive metal.

The Backing Support for the Metals

The backing support for the metals of the folded thermopiles was a cellulose nitrate film formed from a solution of "600-second" cellulose nitrate. The solution is poured on to a glass plate upon which it is spread evenly by a spreader bar or "knife." After the solvent has evaporated, the films on the glass plate are cut into strips 9 cm$\times$1.2 cm and then floated off on water ready to be mounted.

FIG. 1. Strip holder.

[6] H. C. Burger and P. H. Van Cittert, Zeits. f. Physik 66, 210 (1930).

Mounting or Folding the Cellulose Nitrate

In order to obtain a uniform fold of the cellulose nitrate film it is necessary to fold the 9$\times$1.2 cm strip over a "bar," the bar being removed after the fold has set. Various plastic materials were tried for the bar. (The plastics could be removed with solvents to which the cellulose nitrate is inert.) However, better results were obtained by using a Nylon thread held taut by a spring, the thread being withdrawn mechanically after the fold had set.

All subsequent operations are facilitated by mounting the lower ends of the cellulose nitrate folds on a thin strong material. For this purpose we use aluminum strips (0.6 cm wide$\times$22 cm long$\times$0.006 cm thick). The aluminum strips are cut, straightened, and then anodized. The anodized strips are mounted in a holder which keeps them taut. The distance (0.025 cm) from the top edge of the anodized strip to the fold of the cellulose nitrate is controlled by the spacing of the (0.0075 cm diameter) Nylon thread. The adhesion of the cellulose nitrate to the anodized strip is effected by painting "glyptal" solution, three-fourths of the width, along the length of and on both sides of the strip.

The holder (Fig. 1), with the anodized strip and the Nylon thread, is submerged in a vessel filled with water, to the surface of which the

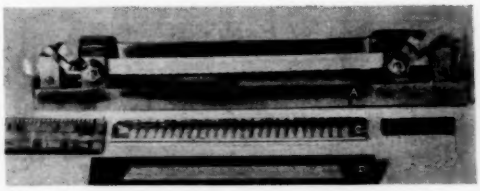


FIG. 2. Evaporation frame.

FIG. 2a. Evaporation frame.

cellulose nitrate film is transferred. The film is oriented so that its long axis coincides with that of the anodized strip. The water is then drained from the vessel so that, as the water surface recedes, the cellulose nitrate film follows and is left in the folded position adhering to the anodized strip. After the film on a strip has dried, the Nylon thread is withdrawn, and the strip with the attached film is ready to be transferred to an evaporation frame.

Frames for Holding Strips during Evaporation

The evaporation frames limit the deposition of each metal to one side of the folded cellulose nitrate strip and to a slight (~0.015 cm) overlap at the fold. Each frame consists of four main parts. One part (*A*) holds the strip taut, yet extendable (by means of a spring); a second part (*B*) serves as a supporting back for the strip; a third part (*C*) is a gold mask (0.005 cm thick) which helps to hold that part of the film which extends beyond the edge of the anodized strip down against the supporting back (*B*) and also confines the metal deposits to bands 0.25 cm wide with a bare space 0.05 cm wide between the bands; the fourth part (*D*) holds the gold mask in place. An evaporation frame is illustrated in Figs. 2 and 2*a*.

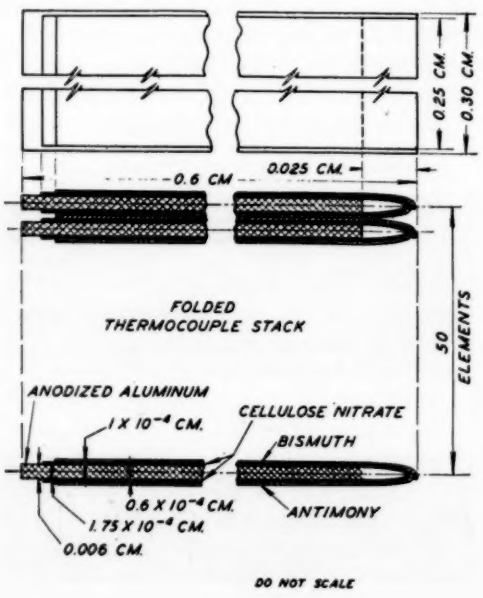


FIG. 3.

After one metal has been deposited, the frames are removed from the bell jar, parts *B*, *C*, and *D* transferred to the opposite side of part *A*, and the second metal is evaporated. Evaporation is accomplished on four strips simultaneously. From each strip twenty-two thermal elements 0.03 cm wide can be cut.

It has been necessary to make trial evaporations for each metal before the metal is deposited on the strips. A cylinder of metal (0.205 g of antimony or 0.445 g of bismuth) is completely evaporated stepwise on to cellulose nitrate films mounted on the mica frames described above. At the end of the evaporation process the films are removed from the bell jar and the electrical resistance of the metal deposit is determined. Trial runs are continued until the deposits have a resistance of 3.8 ohms or less. The frames and strips (illustrated in Figs. 2, 2a) are then placed in the bell jar and the evaporation of the metal carried out.

Stacking and Mounting of Thermopiles

The folded strips are now cut along the bare spaces into individual thermoelements using a cutter similar to a paper cutter. The resistance of each element is measured with a special testing clamp whose jaws grip the element 0.03 cm back from the fold. Those elements having resistances greater than 1.75 ohms are discarded.

The elements are then stacked in units of fifty in a modified press. The folds of the elements are placed against a glass plate for assembly so that all the folds will be in the same plane. Pressure is applied until the electrical resistance is at a minimum. The sides and back of the stack are painted with cellulose nitrate solution while the stack is still clamped, thus cementing the elements into a pile. Figure 3 illustrates one of the elements and also shows how the elements are arranged in a completed thermopile.

The cemented pile is then put into a "skeleton," also a miniature press, and blackened with gold "black." After the skeleton has been inserted into the outer casing of the housing, the piles are ready for use.

TABLE II. Physical dimensions and properties of thermopiles.

	a	*b*	*c*	*d*
Length "hot" to "cold" junction	0.025 cm	0.025 cm	0.035 cm	0.015 cm
Width of each element	0.25 cm	0.25 cm	0.25 cm	0.25 cm
Thickness, antimony	0.6×10^{-4} cm	0.6×10^{-4} cm	0.6×10^{-4} cm	0.6×10^{-4} cm
Thickness, bismuth	1.07×10^{-4} cm	1.07×10^{-4} cm	1.07×10^{-4} cm	1.07×10^{-4} cm
Thickness, cellulose nitrate	1.7×10^{-4} cm	1.7×10^{-4} cm	3.2×10^{-4} cm	3.2×10^{-4} cm
Number of junctions in pile	50	50	50	25
Active receiving area of pile	0.11 cm²	0.11 cm²	0.09 cm²	0.05 cm²
Thermopile resistance	63 ohms	84.5 ohms	56 ohms	32.5 ohms

All the receivers were covered with gold "black." Thermal e.m.f. (measured) Bi-Sb =92 μv/degree Centigrade. *a* and *b* adjacent to one another in the same housing; *c* and *d* adjacent to one another in a second housing.

TABLE III. Response of thermopiles to steady radiation.

	a	*b*	*c*	*d*	*a*+*b* (series)	*c*+*d* (series)	*a*+*b* (parallel)
μv/μ watt/cm²	0.044	0.0513	0.0468	0.0075	0.095	0.054	0.0445
Area (cm²)	0.11	0.11	0.09	0.05	0.22	0.14	0.22
μv/μ watt	0.40	0.467	0.52	0.15	0.431	0.386	0.202

ELECTRICAL CHARACTERISTICS AND SENSITIVITY OF THE FINISHED THERMOPILES

Theory of Thermopile Response

When a thermopile is subjected to interrupted radiation, the a.c. component of the voltage generated by the thermopile may be evaluated as follows:

$$V_{\text{a.c.}} = knm(G/L)\frac{1}{y(f,\tau)}, \qquad (1)$$

where

$V_{\text{a.c.}}$ = the r.m.s. voltage generated at the terminals of the thermopile,
k = a constant of proportionality,
n = the number of "hot" junctions,
m = the volts per degree for each junction,
G = heat absorbed by thermopile receiver (calories/cm²/sec.),
L = heat lost by thermopile (calories/degree/cm²/sec.), and
$y(f, \tau)$ = a functional relation, where f is the frequency of interruption and τ is the relaxation time, or time constant.

The relaxation time may be expressed more explicitly as

$$\tau = \frac{\text{heat capacity of receiver (calories/degree/cm}^2}{\text{heat lost by receiver in unit time (calories/degree/cm}^2\text{/sec.)}}. \qquad (2)$$

In the cases to be discussed here, L is identical with the denominator of (2).

For steady radiation, Eq. (1) reduces to

$$V_{\text{d.c.}} = nm\frac{G}{L}, \qquad (3)$$

where $V_{\text{d.c.}}$ represents the voltage response to steady radiation. Equation (3) permits calculation of the d.c. response from the physical characteristics of the thermopiles.

Physical Characteristics of Thermopiles Tested

We give in Table II the physical and electrical characteristics of four folded bismuth-antimony thermopiles.

Voltage Response to Steady Radiation

Each thermopile was connected to an L. and N., H.S. galvanometer and illuminated with radiation from a "standard" lamp. The responses, corrected for window losses, are given in Table III.

The response to steady radiation is nearly the same for piles *a*, *b*, and *c*, while the response of *d* is much less. This is caused by the fact that *d* has half the junctions of *a*, *b*, and *c*, and the heat loss for *d* is greater than *a*, *b*, and *c*, because of the shorter distance between the hot and cold junctions. The results obtained with the piles in series serve as a check on the accuracy of the measurements. Since each junction acts as its own re-

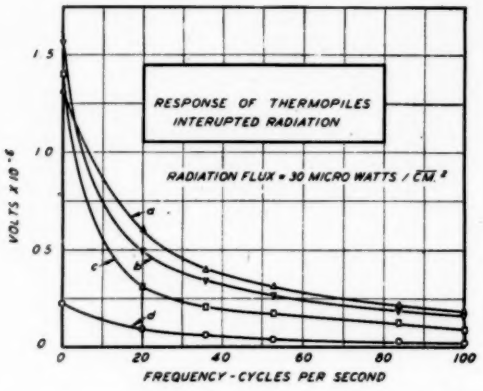


FIG. 4. Response of thermopiles to interrupted radiation. *a*, *b*, *c*, *d*, refer to thermopiles *a*, *b*, *c*, *d*, respectively.

ceiver separately, the voltage output for an extended image is proportional to the number of junctions. The resistance of the pile increases directly with number of junctions.

The heat loss per unit time (L) by the receiver can be evaluated from the measurements in Table III with the aid of Eq. (3). Calculations show that L is approximately 100 times greater for the folded piles than for the single junction evaporated couples operated in vacuum.

Voltage Response to Intermittent Radiation

Radiation from a ribbon filament incandescent lamp was focused on the thermopiles and interrupted by a special rotary shutter. Previous tests showed that this shutter passed radiation of sinusoidally varying intensity. The response of the thermopiles to this radiation at 20, 36, 52.8, 84, and 100 c.p.s. was measured.

The signal from the thermopile passed to a transformer, then to the grid of the first tube of an untuned amplifier. The output from the amplifier was fed to a *GR* wave analyzer. This voltage having been recorded, the thermopile was disconnected from the transformer and a resistor of equal magnitude substituted. A signal from an oscillator, reduced in intensity by 10^6 times, was applied through the resistor, transformer, and amplifier. The output (OP) of the oscillator was adjusted until the same reading was obtained on the wave analyzer as with the thermopile arrangement. The oscillator was then connected directly to the wave analyzer and the voltage output for the setting OP determined. This reading divided by 10^6 equals the output at the terminals of the thermopile. The above procedure was repeated for each frequency. The voltage at zero frequency was determined from the d.c. voltage generated by the thermopile at the maximum transmission of the sector mentioned above. This voltage divided by 2.83 equals the voltage that would be generated at zero frequency. The ratio of the deflections of the galvanometer for the radiation at maximum transmission and that using the standard lamp permitted calculation of the intensity of the intermittent radiation used. The results are plotted in Fig. 4.

Although *a*, *b*, and *c* have nearly the same response at zero frequency, the response to interrupted radiation for *c* and *d* is much less than that for *a* and *b*. This is caused by the much heavier cellulose nitrate backing for *c* and *d*. Thermopile *d* should be faster than *c*, because of the shorter distance between hot and cold junctions. The measurements indicate this.

Speed of response

It is of interest to compare the response observed here with the response of receivers having unique time constants. Since* $v/v_0=(1+2\pi f^2\tau^2)^{-\frac{1}{2}}$ where v = voltage at frequency f, v_0 = voltage at zero frequency for same radiation intensity, f = frequency of interruption of radiation, and τ = time constant, straight lines should be obtained when $(v_0/v)^2$ is plotted against f^2. Figure 5 shows the straight lines obtained for receivers having time-constants of 0.01 sec. and 0.03 sec. The folded thermopiles do not show such a linear relation; therefore they cannot be characterized by a unique time-constant.

A "time-constant" calculated using zero frequency voltages will give an apparently slow response time while a "time-constant" calculated from the high frequency data alone will give an apparently fast response. The "time-constant" is dependent upon the duration of the signal for such a thermopile and therefore does not have the usual significance.

* This relation is valid only for case where heat conduction loss is small compared to radiation loss.

Noise

The "noise" from the thermopiles was less than the "noise" of the amplifier available and could not be measured. The amplifier noise was equivalent to 3×10^{-8} volt.

Discussion of Results

The "folded" thermopiles described here are considerably faster than any thermopiles reported heretofore, yet the response (μ volt/μ watt/cm^2) is greater for all frequencies above 5 cycles per second than for evaporated thermocouples operated in a vacuum.

The calculations for the relaxation time indicate that the use of thinner cellulose nitrate backing material, or another backing material of lower thermal mass, should increase the "speed," and thus the response to interrupted radiation.

The electrical resistance of the individual thermo-elements is $\sim$1.4 ohms whereas the resistance calculated from the resistivity of the metals when evaporated separately is $\sim$0.4 ohms. The source of this additional resistance is not known.

The optimum heat loss for maximum voltage output at different frequencies is not known. Until a satisfactory theory has been evolved, it will be necessary to determine the optimum dimensions empirically. It seems safe to assume that the thermopiles can be substantially improved by further study.

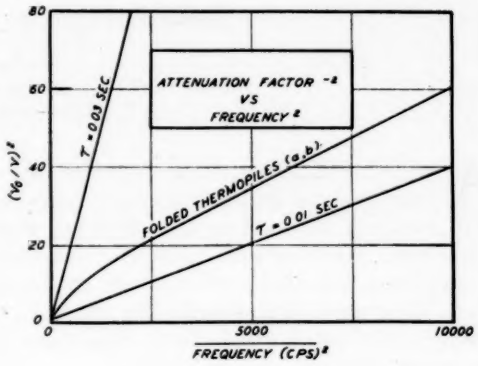


FIG. 5.

Although developed particularly for use as surface piles, the design can be modified easily to form linear thermopiles.

ACKNOWLEDGMENTS

This development is the outgrowth of some ideas held jointly with Mr. Willard E. Buck, who contributed in great measure to the success of the project. The author also acknowledges his thanks to Mr. Alan C. Bemis, Dr. Benjamin M. Siegel, Dr. Walter Stockmayer, Miss Alette C. Curtis, and Mr. James L. Hildebrand.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 36, NUMBER 10 OCTOBER, 1946

Aspheric Corrector Plates for Magnifiers

R. E. HOPKINS
Institute of Optics, University of Rochester, Rochester, New York
(Received August 19, 1946)

A WIDE-FIELD aspheric magnifier, specially designed as a reading glass for persons with impaired vision, has been built at this laboratory. The magnifier consists of a simple planopositive lens plus a thin aspheric corrector plate. This paper will describe the method of designing the system and include a discussion of its performance.

In some optical instruments, like the Schmidt camera for instance, the function and value of the aspheric corrector plate is clear; but in a complicated refracting system like a photographic lens or an eyepiece it is not easy to decide when to use, and how much is gained by using, an aspheric plate. The design of a simple aspheric magnifier provides an excellent example for illustrating a general procedure for studying the function of an aspheric. Third-order theory is used extensively, even in cases where the angles of incidence and obliquities are very large, and where fifth- and high order aberrations are present and large. Before the aspheric is inserted every attempt is made to reduce the effects of high order aberrations in the basic system. The results have been unimpressive when an aspheric plate has been used in a lens system which is "bulging out" with zonal aberrations. It has been found more effective to redesign the system to eliminate zonal aberrations and to insert the aspheric in order to correct the remaining third-order aberrations.

The methods of design and the formulae used are not new. They are a selection of the best methods developed for use with modern calculating machines. Most of the formulae can be derived directly from formulae published by Conrady,[1] and his notation has been used as often as possible.

TYPE OF ASPHERIC PLATE USED

One of the most practical methods of making aspheric surfaces is particularly successful when the aspherics are made in the form of thin corrector plates similar to those used in a Schmidt camera. Thin corrector plates for a 2" focal length $f/0.57$ Schmidt camera can be molded with sufficient quality to give 90 percent of the light in a 30μ central image. These corrector plates may be used in short focal length cameras, condensor lenses, eyepieces, and magnifiers and are the type of aspheric used in the magnifier to be described.

OPTICAL REQUIREMENTS OF A MAGNIFIER

Magnifiers are used in a variety of ways, and their optical requirements differ slightly, depending on their application. Usually, the following characteristics must be considered.

(1) The field should be large. An apparent field of 60° is ideal for most applications. Some magnifiers of 90° have been designed.

(2) The design should be such that the eye can be located at a comfortable distance from the magnifier when viewing the full field.

(3) The lens should be well corrected for spherical aberration over as large a pupil as possible. It is not sufficient to correct a magnifier over a 7- or 8-mm pupil as is done in an eyepiece of a telescope or microscope. In a telescope there is a definite exit pupil, but in a magnifier an observer may view at nearly any distance from the lens and even off the axis. If the lens is well corrected for spherical aberration over the entire aperture, the observer is not confined to a single position for best viewing. This is particularly true for a long focal length reading glass.

(4) For a similar reason the coma should be well corrected.

(5) It is difficult to correct a magnifier for curvature of field; therefore this usually sets the limit to the practical field of view. For short focal length magnifiers the curvature of field becomes very serious, and a 60° field may become impossible to attain.

(6) Astigmatism should be eliminated if possible. In high power magnifiers it is customary to flatten the field artificially by introducing astig-

[1] Conrady's *Applied Optics* (Oxford Press, Oxford, 1929).

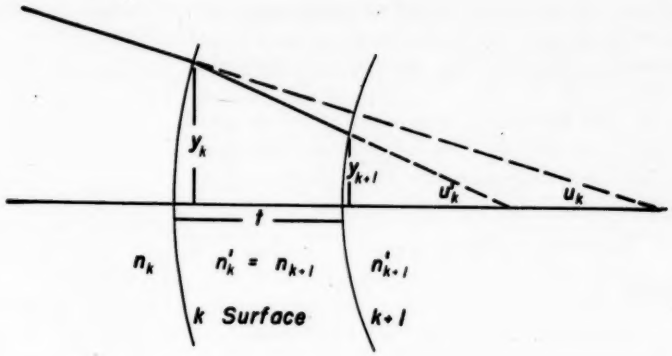


FIG. 1. Diagram illustrating ray tracing data.

matism. However, we prefer a large amount of field curvature (4 to 8 diopters) in preference to introducing as much as 2 diopters of astigmatism. Without the use of a very complex magnifier it is practically impossible to reduce the astigmatism to zero; consequently some compromise has to be made.

(7) The most noticeable monochromatic defect in a magnifier is distortion. In order to avoid criticism of the design, the distortion must be reduced to below 1 percent even at the cost of compromising other aberrations.

(8) Axial transverse color and oblique transverse color must also be well corrected. Two to three minutes of difference in angular position between an object seen in *D* light and *F* light is objectionable. The axial color must again be corrected over as large a pupil as possible or there will be a critical position of viewing required to avoid color fringes.

THE MAGNITUDE OF THE OPTICAL PROBLEM

The design of a magnifier is usually increased in difficulty by the practical requirement that the system must contain as few elements as possible. To attain a large field of view with a long eye distance, it is necessary to have a lens with a large diameter. In order that the magnifier have comfortable eye relief and still cover a 60° field, a lens diameter of about two-thirds the focal length is common. Since the lens should be corrected over the entire aperture, the problem is essentially the same as that of the design of a photographic $f/1.5$ lens which covers a 60° field. Obviously, it is impossible to arrive at the same perfection as a photographic design with the number of elements usually found in a magnifier; but it still remains the goal which the designer must try to attain.

NUMERICAL EXAMPLE OF THE MAGNIFIER DESIGN

In setting up the requirements for the aspheric magnifier which this paper describes, it was considered desirable, for convenience in using the magnifier, that it should cover a field of view at least 5 inches square with the observer's eye approximately 4.5 inches from the lens. A lens of 13.5 cm diameter and 17.78 cm focal length was selected. A completely achromatic lens of this size would be heavy and expensive, but fortunately, for use as a reading glass, it is acceptable to provide an orange filter to partially eliminate the appearance of chromatic defects. Therefore, complete achromatism was not necessary.

Table I shows the calculation page for the third-order aberrations in the design. The aspheric is not included, for it is added at a later stage of the design. The design procedure was as follows.

(a) A paraxial ray entering parallel to the optical axis of the lens was traced by successively applying the following two equations:

$$(n'u')_k = (nu)_k + (n'-n)_k y_k c_k, \tag{1}$$

$$y_{k+1} = y_k - (n'u')_k l_{k,k+1}/n_k'. \tag{2}$$

Figure 1 illustrates the quantities used in these equations.

(b) The focal length and vertex focal length of the lens were then calculated from the formulae

$$f' = y_1/u_0', \tag{3a}$$

$$l_0' = y_0/u_0', \tag{3b}$$

where y_1 is the initial height of the paraxial ray, y_0 is the height of the emerging ray on the last surface, and u_0' is the paraxial ray emerging angle.

(c) The third-order transverse spherical aberration was obtained by adding the individual surface contributions which were calculated from the equation

$$TAC' = yg_4 r(ni)^2(nu/n - ni/n')/2u_0'n_0', \quad (4)$$

where

$$ni = (nyc - nu) \quad \text{and} \quad g_4 = n' - n/rnn'.$$

(d) An oblique paraxial principal ray was then traced through the lens from an entrance pupil -11.605 to the left of the plane surface. The LaGrangian image height was calculated with the equation:

$$n_{l_0}' = (y_0 - l_0'u_0')_{pr}.^2 \quad (5)$$

(e) The axial ray and the oblique ray traces were checked by comparing h_{l_0}' calculated from Eq. (5) and the equation

$$n_{l_0}' = f'u_{1pr}. \quad (6)$$

(f) A constant ρ at each surface was calculated

$$\rho = (ni)_{pr}/(ni). \quad (7)$$

(g) The third-order surface contributions for the kth surface then could be calculated

$$^3\,CC' = TAC'\rho, \quad (8)$$

$$^3\,AC' = CC'\rho/u_0', \quad (9)$$

$$DC' = CC'\rho^2 - \tfrac{1}{2}h'^2\rho g_4 u_0', \quad (10)$$

$$g_4 = (n' - n)c/nn'. \text{ (Petzval curvature).} \quad (11)$$

Summing up the surface contributions gives the third-order monochromatic aberrations (Table I).

TABLE I. Calculation page for third-order aberrations.

c	0	−0.091158		
t		2.7		
n	1	1.617	1	
$-t/n$		−1.66976		
$(n'-n)c$	0	0.056244		
y	1	1		$1_0' =$ 17.7796
nu	0	0	0.056244	$f_0' =$ 17.7796
				$\frac{1}{2}u_0' =$ 8.8898
ni	—	0.147402		
g_4y/c	—	−0.381571		
$(nu/n - ni/n')$	—	0.147402		
TAC'	0	−0.01087		$TA' = -0.01087$
	Paraxial Principal ray			
y	5.802500	6.63738		$H_{1_0}' =$ 8.8899
nu	−0.5	−0.5	−0.126687	
ni_{pr}	−0.5	0.47837		
ρ	—	3.24532		
CC'	—	−0.03528		$Coma' = -0.03528$
AC'	—	−2.0357		$Ast' = -2.0357$
$CC'\rho^2$	—	−0.37157		
$\frac{1}{2}h'^2u_0'\rho g_4$	—	0.25088		
DC'	−0.6861	−0.62245		$Dist' = -1.3086$
g_4	0	0.03478		

RESULTS FROM THIRD-ORDER CALCULATIONS

Transverse Spherical Aberration

The transverse spherical aberration is negative and comes directly from the second surface. The rays outside the extended paraxial ray region will have more negative aberration than pure third order. The third-order constant thus gives an optimistic view of the performance. With an eye movement of one centimeter from the optical center of the lens, an object in the center of the field will appear angularly displaced at least 0.01/17.79 radians or 1.9′.

Coma

The sagittal coma is negative and originates entirely from the second surface; consequently it is to be expected that the actual rays will have more negative coma. The meridional ray coma is -0.11 cm; therefore a one-centimeter movement of the eye in the oblique pencil will cause an object at the edge of the field to appear to move in a sagittal direction through an angle of at least 19′. Both spherical aberration and coma should be reduced if possible.

Field Curvature and Astigmatism

If the astigmatism and coma were equal to zero, the oblique rays would focus upon the Petzval surface which has a curvature equal to the sum of the surface contributions g_4 (Eq. (11)). At the calculated obliquity this would be at a distance of -1.374 cm from the perpendicular plane through the paraxial focus, which corresponds to 0.43 diopter change of focus.

The astigmatism is very large and negative

[2] Subscript pr is used to designate quantities related to the oblique principal ray.

[3] CC' and AC' are the sagittal coma and astigmatism if not subscript with a T.

occurring entirely at the last surface. Again we can expect the higher order aberrations to make the astigmatism even more negative. The sagittal focus will come into focus at -3.41 cm ($1.08D$) and the tangential rays at -7.48 cm ($2.36D$).

Distortion

Since the distortion is negative (pincushion) but well-distributed between the two surfaces, the oblique distortion should not be very different from the third order. The distortion amounts to 14 percent which is very large, since more than 1 percent distortion is usually noticeable.

What Can Be Done with an Aspheric Plate?

In the Schmidt camera the thin aspheric corrector plate is usually located at the center of the aperture stop of the system. The principal ray passes through it at zero height. It is well known that when the plate is moved from the aperture stop the aspheric then contributes to all the third-order aberrations. If the thin aspheric corrector plate adds the amount TAC' of third-order spherical aberration and the plate is located away from the aperture stop, the other third-order contributions are related by the following equations:

$$CC' = TAC'\bar{\rho}, \tag{12}$$

$$AC' = CC'\bar{\rho}/u_0', \tag{13}$$

$$DC' = CC'\bar{\rho}^2 - \tfrac{1}{2}h'^2\bar{\rho}g_4u_0'. \tag{14}$$

These equations are exactly similar to the surface contributions in Eqs. (8)–(11), but in this case $\bar{\rho}$ is the ratio of the height of the principal ray on the aspheric to the height of the axial ray.

In the magnifier problem there are two degrees of freedom with the aspheric. We can choose what TAC' it is to contribute, and then position it anywhere from close to the observer's eye up to the lens or anywhere between the lens and the object plane. There are three interesting cases.

(I) Choose the location of the plate and TAC' to completely eliminate the distortion and the astigmatism. The plate must introduce $DC' = 1.31$ and $AC' = 2.04$. For reasons to be explained later, the aspheric plate will have no power in the paraxial region so that $DC' = CC'\bar{\rho}^2$. By combining Eqs. (13) and (14) we get the equation $DC' = AC'\bar{\rho}u_0'$. By substituting the values required for DC' and AC', we find that $\bar{\rho} = 11.4$, and from Eqs. (12) and (13), we can calculate the required TAC' and the resulting CC'. The summarized results are as follows:

	Aspheric plate	Single lens	Resultant
TAC'	0.00088	−0.011	−0.010
CC'	0.01	−0.035	−0.025
AC'	2.03	−2.04	−0.01
DC'	1.31	−1.31	0

The $\bar{\rho}$ on the plate being 11.4 means that the plate must be located 6.2 from the convex surface of the lens.

(II) By choosing a TAC' and a $\bar{\rho}$ to correct for spherical aberration and distortion, $\bar{\rho}$ comes out to be 4.93. The resultant third-order aberrations are:

	Aspheric plate	Single lens	Resultant
TAC'	0.011	−0.011	0
CC'	0.054	−0.035	0.019
AC'	4.70	−2.04	2.66
DC'	−1.31	1.31	0

(III) If the aspheric plate is placed in contact with the plano side of the lens $\bar{\rho} = -5.8$, and when the TAC' is chosen to eliminate distortion the results are:

	Aspheric plate	Single lens	Resultant
TAC'	0.0067	−0.011	−0.0043
CC'	0.0389	−0.035	+0.0039
AC'	4.01	−2.04	1.97
DC'	1.31	−1.31	0

Case III had enough advantages to warrant completng the design with the aspheric against the positive lens. The practical advantages of having it next to the lens are that the plate can then be cemented, eliminating two glass air surfaces and making mounting very simple. Inspection of the results shows that the aspheric has reduced the spherical by more than one-half. The astigmatism is now positive. The sagittal field is back of the Gaussian image plane by $0.2D$, and the tangential image is back about $1.4D$. The coma has become practically equal to the spherical.

It is possible to make the following estimates on the behavior of the rays out of the near paraxial region. The aspheric plate adds a positive spherical contribution to help annul the negative from the lens, but since the negative contribution is still the larger, and the angles of incidence are

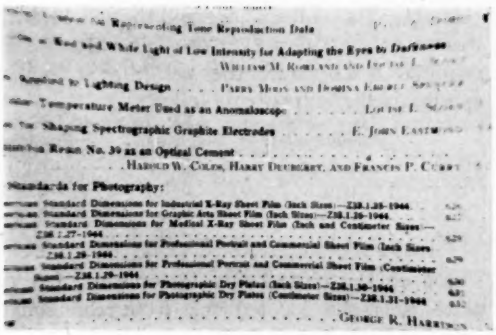


FIG. 2(a). Photograph through magnifier without aspheric.

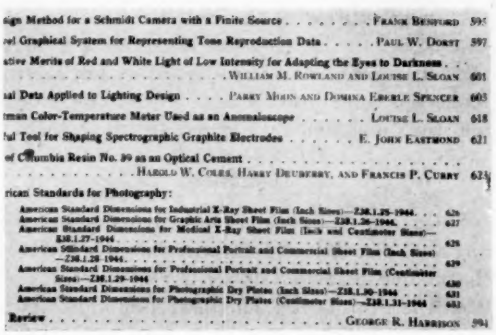


FIG. 2(b). Photograph through magnifier with aspheric.

larger, there will be high order undercorrected spherical aberration. The aspheric plate adds a nearly equal amount of positive coma. The third-order coma cancels out, and higher order coma should tend to cancel out. Likewise, the aspheric adds a large amount of positive astigmatism, and the actual astigmatism will be slightly more positive than calculated from third order. The largest distortion contribution now is caused by the aspheric indicating that there will be a positive zone in distortion. By correcting for zero distortion at the edge of the field, there will be a slight amount of negative distortion at the half field. Calculations show that with -0.36 percent distortion for 26° obliquity there is -0.7 percent at 15°.

Method of Calculating the Aspheric Plate

The determination of the aspheric was very simple. An oblique principal ray at 26° was traced up to the position of the aspheric surface. The slope was then adjusted until the distortion was reduced to 0.36 percent. A fourth-order curve was then put through the point with this known slope. The resulting equation was $x=-0.0001014y^4$ (with dimensions of x and y in cm).

Since the aspheric plate has negative power near the edge, it was made out of very dense flint ($N_D=1.72$, $V=29.3$) in order to correct as much transverse color as possible.[4] The lens without the aspheric plate had 19 minutes of transverse color at a 26° obliquity. With the aspheric the color was reduced to 14 minutes.

Photographs of Results

The photographs shown in Figs. 2(a) and (b) were taken through the magnifier with and without the aspheric plate. The 4-mm diameter camera aperture was located 11.7 cm from the plane surface of the lens for both photographs. They were taken on Process Panchromatic film with a No. 22 filter. Since the camera lens focal length was 10.5 cm and the diagonal of the picture was 9.5 cm, the photographs show a full field of 48.6°. The lens, however, was designed to cover 52° at this eye relief.

The photographs clearly show how the aspheric completely eliminates the distortion and greatly improves the astigmatism. The effects of spherical aberration and coma, however, cannot be seen in the photographs. Coma and spherical aberration merely affect the apparent position of the field of view as the observer moves around in the eye pupil. A movement of 1 cm from the optical axis causes the image to move angularly 1.9′. This amount of movement is not objectional if the magnifier is used monocularly. However, observers try to use it binocularly on account of its large diameter. If the observer has an interpupilary distance of 64 mm and the lens is adjusted to collimate the paraxial rays, there will be at least a 2° divergence between the lines of sight for the two eyes. The present magnifier thus cannot be used binocularly.

[4] The use of a highly dispersive correction plate to reduce residual chromatic errors and particularly lateral color was first suggested to me by Dr. Brian O'Brien. This principle can be used in a number of optical instruments (wide angle eyepieces, for example).

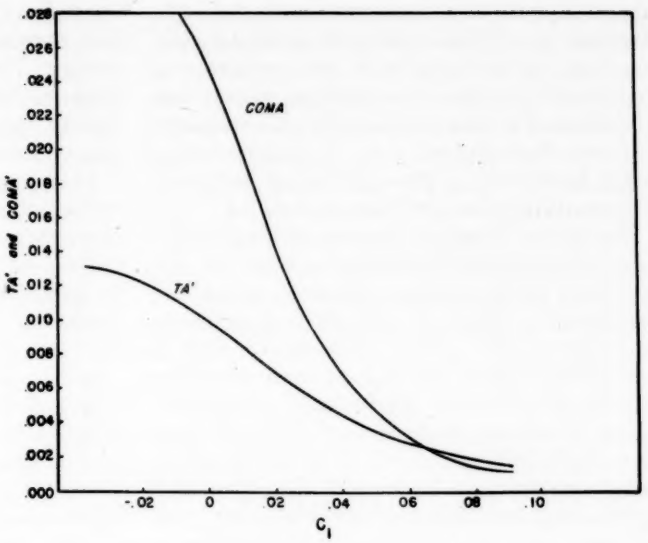


FIG. 3. Variation of TA' and $Coma'$ with bending. TA' and $Coma'$ shown are total from lens and aspheric. At each bending the aspheric is adjusted to correct for distortion and astigmatism.

FURTHER IMPROVED DESIGNS

The planoconvex lens has the practical advantage that the thin corrector plate can be cemented on the plano side. However, by using the extra degree of freedom offered by bending the lens, interesting and more satisfactory results can be obtained theoretically. To investigate the possibilities the lens was bent into a number of different shapes and all the third-order errors calculated. Then the $\bar{p}$ and TAC' of the aspheric were calculated for each case to make the distortion and astigmatism zero. In Fig. 3 the aberrations for lens and aspheric are plotted against the curvature of the first surface of the lens. At the bending $c_1=0.085$, the third-order results are excellent. The spherical aberration and coma are less than half what they are in the planoconvex lens with the aspheric in contact. The spherical aberration is distributed well between the surfaces; therefore the high order spherical should be small. The astigmatism and distortion are zero. The individual surface contributions for this lens are as follows:

	Surface 1	Surface 2	Aspheric	Total
TAC'	−0.00130	−0.00111	0.000749	−0.0016
CC'	−0.01520	0.00243	0.01177	−0.0010
AC'	−3.1898	−0.09563	3.2854	0
DC'	−2.9262	0.023193	2.9029	0

Inspection of contributions on the first surface indicate the difficulty that would be encountered with this lens. The astigmatic and distortion contributions are so enormous that a large negative zone in both these aberrations can be expected. This means that the aspheric plate would have to give a larger positive distortion contribution than expected, and in turn it would not cancel out the other aberrations as well.

To eliminate the effects of the large negative contribution on the front surface, a second plate of sixth or higher order could be placed in front of the offending surface. This plate could be adjusted to eliminate the high order distortion from the first surface. We could then insert the first aspheric to correct for the third-order distortion and astigmatism, and the complete system should perform as expected.

It was doubtful if the added complexity was worth while in our particular problem with a 7″ focal length reading glass; so the design was not completed. This additional aspheric may be, however, the most practical solution for a lens which is to be used as a two-eye magnifier where it is essential to have the spherical aberration reduced to a very small value.

OTHER APPLICATIONS

The same principles have been applied in the design of achromatic magnifiers. A 7″ focal length achromatic reading glass was also designed. In

this design the most favorable position for the aspheric was midway between the focal plane and the lens. As a reading glass this position was unfavorable because of the difficulty of mounting the plate. The design results on this magnifier were excellent, but since the single unachromatized lens with a filter performed well, the achromatic lens has not been constructed.

As the focal length of the magnifier is reduced, the field curvature becomes serious. At the obliquity U_{pr} the diopters of field curvature can be calculated from the following approximate formula: $D=100 \tan^2 U_{pr}/3f'$ (f' is expressed in cm). If the focal length of the magnifier is to be equal to 2.5 cm (10×), the field will be curved by about 4.5 diopters at the edge of a 60° field. This amount of field curvature is too large; therefore some attempt has to be made to flatten the field to less than 0.5 diopter of field curvature. The most efficient way to do this is to add a negative field flattener as close to the focal plane as possible. The lens can then be designed in the same manner as described above, and the aspheric plate can be adjusted to correct for astigmatism and distortion simultaneously.

An aspheric can be used in an eyepiece in a similar manner. Simultaneous adjustment of distortion and astigmatism is accomplished by utilizing an aspheric plate. On the other hand, the spherical aberration and coma are usually no problem.

ACKNOWLEDGMENTS

Most of the numerical calculations in the above design were made by Miss Louise Rivoli, and several members of the laboratory assisted in making the lens and the aspheric corrector plate. Their aid and cooperation are greatly appreciated.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 36, NUMBER 10 OCTOBER, 1946

A Screening Test for Defective Red-Green Vision*

LEGRAND H. HARDY, GERTRUDE RAND, AND M. CATHERINE RITTLER
Knapp Memorial Laboratories, Institute of Ophthalmology, College of Physicians and Surgeons of Columbia University, New York, New York

(Received July 27, 1946)

PROBABLY the demand most frequently made of those who have had experience in testing for color-blindness is for a simple means of detecting the presence of defective color vision—in other words, for a simple screening test adequate to differentiate between normal and defective color vision. The demand comes from the military services, from ophthalmologists, from industry with its increasing use of color, and from vocational high schools where too often a boy is trained for a trade only to find near the end of his training that he can not meet its color vision requirements. Tests have been produced which, when properly administered and the critical scores established, will meet this need, as was shown by us in previous reports[1]; but these tests are of foreign make and not now available in quantity. Therefore, their evaluation has at the present time more theoretical than practical value. Other tests are under production which are planned both to detect the presence of color deficiency and to analyze its type and amount. But lag in production affects the making of color tests as well as of automobiles, and it may be some years before these tests can be put into general use. In the meantime something must be found to serve as an adequate screening test of color deficiency—something which is immediately available.

The only available test which gives promise of success for this purpose is the Pseudo-isochromatic Plates for Testing Color Perception compiled in 1940 by the American Optical Company

* Based on 18 Pseudo-Isochromatic Plates from the American Optical Company's Compilation.

[1] L. H. Hardy, G. Rand, and M. C. Rittler, "Tests for the detection and analysis of color-blindness. I. The Ishihara test: an evaluation," J. Opt. Soc. Am. **35**, 268 (1945); "II. A comparison of editions," *ibid.* **35**, 350 (1945); "III. The Rabkin Test," *ibid.* **35**, 481 (1945).

for the military services. This compilation consists of 46 plates copied from either Ishihara's Test for Color Blindness or Stilling's Pseudo-isochromatic Plates for Testing the Color Sense. Designed principally for the detection of deficiencies in red-green perception, it has been widely used but with generally unsatisfactory results. In part this is because of inaccurate color reproduction and a poor selection of plates for the compilation; in part because no critical scores are provided. That is, no information is given by the compilers of the test as to the number of plates likely to be failed by those whose color perception is normal or of the number likely to be passed by those whose color perception is deficient.

In this report of our experience with the compilation administered as one of a battery of color tests given to a large group of subjects having normal, low normal, or defective color vision of varying degrees of defect, we plan to show that this compilation of plates, even when properly administered under the quality of illumination accepted as standard for color testing,[2] is inadequate to detect with reasonable accuracy the presence of color deficiency. On the constructive side, we shall present a selection of 18 plates of this compilation which when properly administered *is* adequate to screen these with even minor degrees of defective red-green vision from those whose color vision is within normal limits.

As previously reported, our general procedure was to administer a large battery of color tests, both laboratory and clinical in type. On the basis of the composite picture obtained, we classified the color vision of each subject into the following categories: normal, low normal, low discrimination sufficient to amount to a color defect, anomalous trichromasy (protanomalous, deuteranomalous, tritanomalous), and dichromasy (protanopia, deuteranopia, tritanopia[3]). A description of the color vision typified by these terms has already been reported.[4] Our data thus permit us to evaluate each test of the battery as a diagnostic or screening medium, and as a medium for the differential classification of the type and extent of the color vision defect.

It is not our intention to present here a detailed analysis of the responses to each plate of the American Optical Company's compilation made by each of the types of defective color vision, as we did for the Ishihara and Rabkin tests. There are no specific plates in this compilation designed to analyze type of defect. We plan merely to show the distribution of errors made on the compilation for 365 subjects whom we judged on the basis of our battery of tests to have normal color vision and for 117 subjects similarly judged to have defective red-green vision. Our method of determining whether the color vision fell within normal limits has already been described.[5] The subjects tested were not selected on the basis of any statistical survey. They included patients referred to us by ophthalmologists, men rejected by the Armed Forces because of defective color vision, and subjects obtained from some of the New York City high schools and from Vanderbilt Clinic. Only the 40 plates designed to test literate subjects were scored. The demonstration plates (25, 26, 46) and the "path" plates for illiterates (35, 36, 45) were not included. For all testing the book was placed on the rack of the Macbeth Easel Lamp. This equipment, which was designed for color-blindness testing, provides 45 footcandles of illumination the quality of which is the closest approximation to that of ICI Illuminant *C* commercially available.[2] The light fell on the test. at an angle of 45 degrees, the viewing angle was normal or 90 degrees, and the testing distance was approximately 30 inches.

THE 40-PLATE SERIES

Results for the 40-plate series are given in Fig. 1. The data for the 365 subjects judged as having normal color vision are presented at the left of the chart as rectangles in outline; for the 117 subjects with defective color vision, at the right of the chart as black rectangles. Number of plates failed (error score) in increasing groups of

[2] L. H. Hardy, "Standard illuminants in relation to color-testing procedures," Arch. Ophth. **34**, 278 (1945).

[3] We have had the opportunity to study too few cases of tritanomaly and tritanopia to warrant their inclusion in the present study. The experimental subjects reported here have predominantly defective red-green vision.

[4] L. H. Hardy, G. Rand, and M. C. Rittler, "Color vision and recent developments in color vision testing," Arch. Ophth. **35**, 603 (1946).

[5] L. H. Hardy, G. Rand, and M. C. Rittler, "Tests for the detection and analysis of color-blindness. III. The Rabkin test," J. Opt. Soc. Am. **35**, 484 (1945).

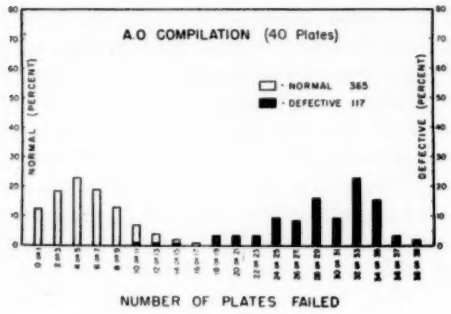


FIG. 1.

2 is plotted on the horizontal coordinate and percentage of cases failing each group on the vertical coordinate.

Normal Color Vision

As can be seen in Fig. 1, only 12.5 percent of those with normal color vision failed none of the plates or only 1; 46 percent failed 6 or more plates; the maximum number failed was 17 plates. Rarely was there complete failure to see a digit as occurs frequently when color vision is defective. Rather, the usual type of error was a misreading of digits because of lack of careful scrutiny, as 48 for 43 (Plate 2); a misreading of Stilling digits of German formation, as 25 for 75 (Plate 30); the reading of the digits supposed to be invisible to those with normal color sense (Plates 33 and 34); or the partial misreading of a digit of the Ishihara plates, as 20 for 29 (Plate 11). It should be noted that the printing of those sections of the digit supposed to be seen only when red-green perception is defective, appears in stronger coloration than in the original plates. This is particularly marked with the green-blue and blue portions of these digits. There is, therefore, a greater incentive for the normal subject to read them, or parts of them, as reproduced by the American Optical Company than as presented in the original Ishihara test.

It may seem strange to one who has not had a wide experience in administering this compilation to a large unselected group varying in intelligence and interest that so many errors of these types can be made by persons with normal color vision. This, however, has been the general experience of those who have administered it in the military services, in industry, and in schools. The compilation is definitely faulty in its inclusion of so many plates subject to misreading because of other factors than defective color vision.

Defective Color Vision

As can be seen in Fig. 1, subjects whose color vision was diagnosed as defective on the basis of our battery of tests, failed as few as 10 and as many as 39 of the 40 plates used.

In brief, there is a range of 10 to 17 plates failed which includes the error scores of some whose color vision is normal and some whose color vision is defective. Diagnosis on this compilation is therefore not possible when error scores occur within this range. In other words, no critical score can be established for this compilation which will differentiate between normal and defective color performance.

THE 18-PLATE SERIES

The problem, however, cannot be abandoned at this point. This test must if at all possible be made to do the work since, as stated earlier, it is the only one worth considering which is commercially available at the present time. We feel the problem can be met fairly satisfactorily by a selection of 18 plates from the American Optical Company's compilation. In selecting each plate we were not guided solely by statistical criteria. Our aim was simply to select a series of plates that would yield a critical score between normal and defective red-green vision. In general, we discarded as far as possible those plates on which insignificant errors were made because of other factors than defective color perception. At the same time we needed to retain some plates which were failed by those whose color defect was mild in degree even if these were occasionally misread by members of the normal group. In other words, we could not sacrifice the sensitivity of the series to detect minor degrees of color deficiency in the interest of securing a perfect response from all whose color vision is within normal limits. Also we wished to include some plates failed by those whose color defect was marked. These considerations barred us from searching for plates passed by all the normal group and failed by all the color defective group.

Further, attention had to be given to excluding plates which vary markedly from copy to copy of the compilation. For this reason Plates 39, 40, 43, and 44 were rejected from our selection. Fundamentally, the selection was a trial and error procedure, being adjusted and readjusted according to the responses given chiefly by those normal subjects who failed many plates and by those color defective subjects who failed few plates.

Our final selection includes Plates 3, 4, 5, 6, 8, 9, 12, 13, 16, 17, 19, 20, 21, 23, 27, 29, 41, and 42. The results are shown in Fig. 2 for 365 subjects whose color vision we judged to be within normal limits and 150 whose color vision we judged to be defective. As in Fig. 1, the data for the normal group are presented at the left of the chart as rectangles in outline; for the color defective group, at the right as black rectangles. Number of plates failed (error score) is plotted on the horizontal coordinate and percentage of cases failing each successive number of plates on the vertical coordinate.

Normal Color Vision

As can be seen in Fig. 2, 93 percent of those we judged on the basis of our battery of color tests to have normal color vision failed none of the 18 plates or only 1; 4 percent failed 2 plates; 2 percent failed 3 plates; and 1 percent failed 4 plates.

Defective Color Vision

All those we judged on the basis of our battery of color tests to have defective red-green vision failed 5 or more of the 18 plates; 2 percent failed all the plates; the number of plates failed most frequently was 15 and 16; 80 percent failed 11 or more plates. No dichromats in our group failed fewer than 12 plates, and anomalous trichromats whose defect was medium or marked failed from 9 to 17 plates. Thus error scores can not be considered as indicating degree or extent of color deficiency except within very rough limits.

The value of this 18-plate selection has been tested by us for two years in schools and other surveys, also by a number of ophthalmologists connected with testing for civil aviation. It has rendered good service and so far has given satisfactory information. We feel confident in recommending this selection to meet the present need for a test adequate to separate those with normal from those with defective red-green vision. This recommendation is contingent on the test being administered under the standard source of illumination.

The critical importance of illumination can not be too strongly stressed in the use of any test designed to screen out those with defective color vision. Just "natural daylight" will not do. It is far too variable in quantity and quality to provide a standard test situation. For example, if the amount of light is too low, the error scores of some with normal color vision will lie above the critical borderline. If the light is too blue, as when reflected from north sky or from snow, again some normal subjects will have an incorrectly high error score. This is mainly because the bluish digits of the Ishihara plates (those usually seen only when red-green vision is deficient) are enhanced in coloration and are, therefore, often responded to in whole or in part by the normal subject. On the other hand, if the illumination is too yellow, as in direct sunlight or illumination by incandescent lamps, the reverse situation holds. That is, these bluish digits are rendered less pronounced, sufficiently so that many whose color deficiency is mild give the correct response and thus lower their error score sufficiently to cross the boundary line into the region of normal error score. The effect of this last type of faulty illumination has already been reported.[6]

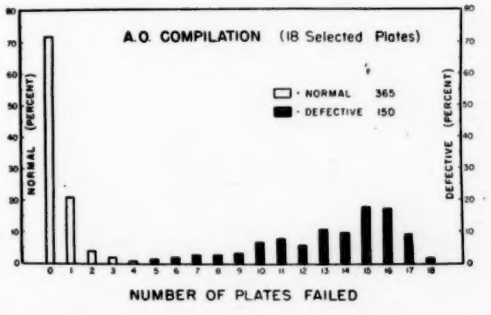


FIG. 2.

[6] L. H. Hardy, G. Rand, and M. C. Rittler, "The effect of quality of illumination on the results of the Ishihara test," J. Opt. Soc. Am. **36**, 86 (1946).

In conclusion we wish to repeat that the 18 plate selection is intended for screening purposes only. It is neither qualitative nor quantitative—that is, it does not classify type of red-green deficiency or, excepting for the extremes of error score, degree of deficiency. It should be pointed out again that no provision is made for detecting the rare forms of blue-yellow deficiency.

Appended is a copy of the instruction sheet we have prepared for those who may be interested in utilizing this 18-plate selection as a screening test for red-green defective vision.

Instructions for Utilizing 18 of the American Optical Company's "Pseudo-isochromatic Plates for Testing Color Perception" as a Screening Test for Defective Red-Green Vision.

LeGrand H. Hardy, Gertrude Rand, and M. Catherine Rittler

Source of Illumination: The Macbeth Easel Lamp. This can be purchased from Macbeth Daylighting Company, 227 W. 17th Street, New York 11, New York, for $20.

Test Material: A selected series from "Pseudo-isochromatic Plates for Testing Color Perception," American Optical Company, 1940. The following plates are used: No. 25 for demonstration; Nos. 3, 4, 5, 6, 8, 9, 12, 13, 16, 17, 19, 20, 21, 23, 27, 29, 41, and 42 for testing. Other plates are to be covered with a sheet of black paper.

Instructions: The subject should be seated with his eyes about 30 inches from the test book in position on the rack of the lamp. He is shown Plate No. 25 (a red 12 on a blue background, a malingering test) and instructed to read the number or numbers on this and the following charts. The examiner allows for response about 2 sec. per plate. No further instructions are to be given and no questions asked. The subject is not allowed to trace the patterns or touch the test plates. Responses should be immediate (within 2 sec.) Hesitant or studied responses are viewed with suspicion. Responses are to be entered on a score sheet similar to that given below. Number of plates failed is the subject's error score.

SCORE SHEET Name_____ Date_____

PLATE NO.	NORMAL RESPONSE	SUBJECT'S RESPONSE	PLATE NO.	NORMAL RESPONSE	SUBJECT'S RESPONSE
3	56	——	17	25	——
4	27	——	19	5	——
5	8	——	20	3	——
6	6	——	21	97	——
8	42	——	23	56	——
9	56	——	27	89	——
12	57	——	29	86	——
13	86	——	41	15	——
16	9	——	42	74	——
					Total Errors___

Interpretation of error score: Incorrect response to 4 or less plates indicates normal color vision; incorrect response to 5 or more plates indicates defective red-green vision.

Limitations of test: This interpretation of error score holds only when the test is administered under the standard source of illumination. The test is for screening purposes only. It is neither qualitative nor quantitative—that is, it does not classify type of red-green defect or amount of defect. No provision is made for detecting defective blue-yellow vision.

Knapp Memorial Laboratory of Physiological Optics
Institute of Ophthalmology
New York 32, New York

Letters to the Editor

"Axial Image-Inversion"

GORDON L. WALLS
Department of Psychology, University of Rochester, Rochester, New York
September 20, 1946

KIRKPATRICK[1] has recently suggested that a telescope inverts the image axially ("fore and aft") as well as vertically, for he finds that through a galvanometer-type telescope the projections on a rough wall may appear as indentations. Hudson[2] reports similar phenomena and raises the question whether the inversion is objective or subjective. He offers two photographs, made through optical systems, which show apparent convexities. Upon inverting one of these, the convexities are readily seen as the concavities they really are, leading Hudson to think the axial "inversion" subjective. On inversion of the other photograph, however, the apparent convexities tend to persist as convexities—leaving Hudson again doubtful of the subjectivity of the phenomenon.

It is difficult to see how any physicist could think for a moment that the axial "inversion" is objective. If it were, all the rules for focusing would be topsy-turvey. We should have had to consider, all these years, that the observed axial chromatic aberration of a lens is anomalous, for if the inversion were real the red focus should lie between the lens and the violet focus.

The phenomenon under discussion is familiar to psychologists as the "cameo-intaglio illusion." It has its whole basis in the fact that the natural illumination of objects is from above downward so that shadows usually lie *beneath* the objects that make them. If such an object as a hemispherical rivet head on the side of a boiler is illuminated from above, its lower part will be shadowed. If it were a hemispherical dent, its upper part would instead be shadowed (by the upper part of its rim). A photograph of such rivet heads, inverted, presents them as dents. The telescopic image of a top-lighted convexity may similarly present it to visual perception as a concavity, if the telescope is non-erecting.

Whether top-lighted convexities in an inverted image or photograph will be seen as concavities or not depends upon the observer's familiarity with the real object and related objects. If the object is totally unfamiliar, the inversion in perception is ordinarily easy. If it seems to be recognizable and describable (whether correctly or falsely—it does not matter), it will be capable of perceptual axial inversion provided this is imaginable for the object as the observer conceives it to be. The boiler rivet head is inversable because the observer is able to accept the idea that a boiler wall can have a dent in it. If the object is very familiar, entirely devoid of ambiguity, incapable of existing in both convex and concave versions, then it can be perceived in only one way. A photograph of a human face will refuse to "invert" axially when turned upside down, no matter how the camera subject was lighted, because the observer's mind would reject the idea that any face could embody a concave nose, brow, and cheeks.

The rules are much the same as those which govern whether a binocularly fused stereo-pair of photographs will yield a pseudoscopic percept or a persistently stereoscopic one when the picture appropriate to the left eye is haploscopically presented to the right eye and vice versa. It is impossible to perceive a human face pseudoscopically—its true conformation is too familiar to all observers.

[1] P. Kirkpatrick, Am. J. Phys. **13**, 203 (1945).
[2] D. R. Hudson, J. Opt. Soc. Am. **36**, 492 (1946).

A Simple Cell for Spectroscopic Examination of Volatile Liquids

ROBERT A. OETJEN,* WILLIAM M. WARD, AND JUNE A. ROBINSON
Beacon Research Laboratory of The Texas Company, Beacon, New York
August 28, 1946

IT is the purpose of this note to describe a cell which has been found satisfactory for determining infra-red absorption spectra of liquids which have low boiling points. The cell is of simple construction and is easily dismantled for cleaning. Although the cell described here has been used for infra-red investigations, it is clear that, by substituting appropriate materials in its manufacture, it can be used in conducting investigations in other spectral regions.

Figure 1 shows the construction of the cell. The two potassium bromide windows are separated by a platinum spacer of the desired thickness which is shaped so that it forms a closed rectangle into which the sample may be introduced. Two clamping plates, made of steel, are employed for holding together the "sandwich" formed by the

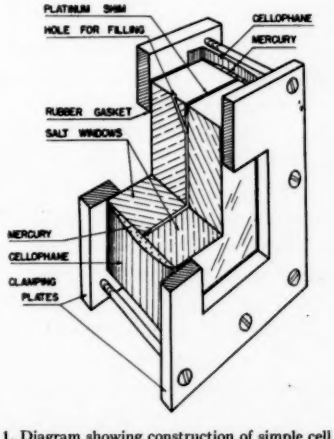


FIG. 1. Diagram showing construction of simple cell for spectroscopic examination of volatile liquids.

windows and the platinum shim. To protect the windows from damage by the clamping plates, rubber gaskets are cemented to the portions of the clamping plates which would otherwise be in contact with the windows. A piece of Cellophane is folded so that it fits around two sides and the bottom of the sandwich comprising the cell windows and the shim. The edges of this Cellophane are folded so that they lie against the outer surfaces of the cell windows, and are clamped between the rubber gaskets and the cell windows during the assembly of the cell. After the clamping screws have been tightened to provide fairly uniform pressure all around the Cellophane mercury is poured into the space between the Cellophane and the edges of the windows and shim. Sufficient mercury is used to bring its level to the top of the windows. The sample is introduced into the cell by means of a hypodermic syringe whose needle is inserted into one of the two filling holes drilled diagonally through the top of one of the cell windows with a number 53 drill. After filling, a small piece of synthetic rubber, or other plastic material which is not affected by the sample, is used to stopper the hole, and additional mercury is poured onto the top of the cell until the stopper is completely covered. In order to keep the mercury on top of the cell, the Cellophane is cut long enough to extend at least one-half inch above the windows. When the cell is assembled, it is necessary for the liquid to pass through the mercury in order to escape. To assure that the cell will remain full in the event that some of the sample is lost during filling or seeps out past the shim to the mercury, a reservoir holding between 0.1 and 0.2 ml is provided. This is made by milling a channel into the salt at the end of one of the filling holes. The milling is done in a drill press by letting the rotating drill dig very lightly into the salt which is moved back and forth on the bed of the drill press. In such a cell, diethyl ether has remained overnight in the area of the cell through which radiation passes.

For the cells used in this laboratory, where each window has dimensions about $40\times65\times10$ mm, it has been found that Cellophane 0.075 mm thick is ample to hold the weight of the mercury required when the Cellophane "fits" reasonably well. Difficulty of assembly is minimized if the Cellophane is folded and the corners cemented so that it has the approximate shape of the edges of the unit consisting of the windows and shim.

If the thickness of the cell is to be assumed to be that of the shim, it is necessary that the windows be flat almost to the extreme edges, and that the shim also be flat. If there is doubt as to the exact thickness, it may be determined before filling the cell, through an interference method such as the one discussed by Smith and Miller.[1]

The cell is dismantled by pouring out the mercury and removing the clamping screws. Each part may then be cleaned as thoroughly as desired.

* Now at Mendenhall Laboratory of Physics, The Ohio State University, Columbus, Ohio.

[1] D. C. Smith and E. C. Miller, J. Opt. Soc. Am. **34**, 130 (1944).